RESEARCH ON GaAs QUANTUM-COUPLED STRUCTURES THAT CAN BE USED AS ELECTRON DEVICES

Texas Instruments Incorporated
13500 North Central Expressway
P. O. Box 655936, M.S. 105

26 January 1989

Final Technical Report, 16 June 1987 - 31 December 1988

Prepared for

Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217
Dallas, Texas 75265

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT This document has been approved for public release and sale; its distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) 08-89-54	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Texas Instruments Incorporated 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and Zip Code) 13500 N. Central Expressway Dallas, Texas 75265	7b. ADDRESS (City, State, and Zip Code)
8a. NAME OF FUNDING/SPONSORING ORG. Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-87-C-0700
8c. ADDRESS (City, State, and Zip Code) 800 North Quincy Street Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. 414e200 ---02 / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Research on GaAs Quantum-Coupled Structures That Can be Used as Electron Devices

12. PERSONAL AUTHOR(S)
R. T. Bate, W. R. Frensley, M. A. Reed

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Technical	16 June 1987-31 Dec 1988	26 January 1989	147

16. SUPPLEMENTARY NOTATION
Reproduction in whole, or in part, is permitted for any purpose of the United States Government.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The objective of this program was to conduct a theoretical and experimental investigation of the physics of quantum-coupled semiconductor structures, with the ultimate goal of developing an integrated circuit technology based on quantum-coupled devices. The prototype structure on which both the theoretical and experimental efforts were based was the quantum-well resonant-tunneling diode. The theoretical work answered questions concerning the dc and small-signal ac behavior of this device and some very fundamental questions concerning the theory of open quantum systems. The theory of open

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☑ UNCLASSIFIED/UNLIMITED ☑ SAME AS RPT. ☐ DTIC USERS	UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Larry R. Cooper	22b. TELEPHONE (Include Area Code) (202) 696-4619	22c. OFFICE SYMBOL 1114SS

DD FORM 1473 84 MAR

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. ABSTRACT (Continued)

systems is a necessary part of the development of quantum device technology because all electron devices are open systems.

The experimental part of the program focused on two areas of investigation. The first concerned the behavior of quantum dot structures, which are resonant-tunneling diodes whose lateral dimensions (perpendicular to the direction of current flow) have been reduced to quantum scales by microlithography and (at present) etching. We have successfully explained the spectroscopy of these structures as revealed in their I-V characteristics. The second area of investigation compared the I-V curves of structurally well-characterized resonant-tunneling diodes to theoretical models. We found that the resonant peak voltages sensitively depended on the precise details of the epitaxial structure. The fitting of the theoretical model to the experimental I-V curve appears to be a more precise way to determine these details than any existing direct technique.

TABLE OF CONTENTS

LIST OF APPENDIXES

FINAL TECHNICAL REPORT
FOR CONTRACT NO. N00014-87-C-0700
RESEARCH ON GaAs QUANTUM-COUPLED STRUCTURES THAT CAN BE USED AS ELECTRON DEVICES

I. INTRODUCTION

The objective of this program was to conduct a theoretical and experimental investigation of the physics of quantum-coupled semiconductor structures, with the ultimate goal of developing an integrated circuit technology based on quantum-coupled devices. The prototype structure on which both the theoretical and experimental efforts were based was the quantum-well resonant-tunneling diode. The theoretical work answered questions concerning the dc and small-signal ac behavior of this device and some very fundamental questions concerning the theory of open quantum systems. The theory of open systems is a necessary part of the development of quantum device technology because all electron devices are open systems.

The experimental part of the program focused on two areas of investigation. The first concerned the behavior of quantum dot structures, which are resonant-tunneling diodes whose lateral dimensions (perpendicular to the direction of current flow) have been reduced to quantum scales by microlithography and (at present) etching. We have successfully explained the spectroscopy of these structures as revealed in their I-V characteristics. The second area of investigation compared the I-V curves of structurally well-characterized resonant-tunneling diodes to theoretical models. We found that the resonant peak voltages sensitively depended on the precise details of the epitaxial structure. The fitting of the theoretical model to the experimental I-V curve appears to be a more precise way to determine these details than any existing direct technique.

II. THEORY

The theoretical investigations conducted as part of this contract involved the extension and elaboration of the quantum kinetic transport theory developed under the previous program, Contract No. N00014-84-C-0125, "Research on GaAs Quantum-Coupled Structures That Can Be Used as Electron Devices." This theory assumes that a system exhibiting quantum electron transport is a finite open system, and it represents the state of this system by the Wigner distribution function. The openness of the system is modeled by the boundary conditions applied to the Wigner function as it evolves in time as prescribed by the Liouville equation. This theory has been applied principally to the study of the quantum-well resonant-tunneling diode (RTD).

We found that the agreement between the I-V curves obtained from the quantum kinetic theory and the more conventional stationary-state scattering theory is considerably better than our earlier results indicated. The source of the earlier discrepancy was an error in the evaluation of the scattering-theory current density. In the course of adapting the scattering state program code to the needs of another research program, we discovered a coding error in the kinematic factor that contributes to the current density. The correction of this error not only improves the agreement between the scattering and quantum kinetic theories, but also improves the agreement between the scattering theory and experiment for resonant-tunneling diodes with AlGaAs barriers in the direct energy gap range. (However, a significant discrepancy remains between theory and experiment for devices with AlAs barriers.) The new comparison is displayed and discussed in the manuscript included as Appendix A.

Another aspect of the quantum kinetic theory that has been further developed during this program is its use to evaluate the small-signal ac response of tunneling diodes. Appendix B describes the results of this task in detail. This work is significant because it provides an explicit prediction of the linear and nonlinear responses of the RTD and displays the differences between linear and nonlinear behavior. A particularly interesting point is that the analysis demonstrates the nonlocality of the current response as the frequency is increased and clearly shows the

changeover from the electronic to the optical regime (see Figure 5 of Appendix B).

The effects of inelastic collision processes are an essential element in the description of classical semiconductor devices and are expected to be significant in quantum devices as well. As the first step in modeling such effects, we added a classical Boltzmann collision operator to the Liouville equation for the Wigner function. The results are presented in Appendix C, which shows that the quantitative effects of phonon scattering on the I-V characteristics of the RTD are quite modest. While our present analysis does not include the possible scattering mechanisms and treats those included only semiclassically, we believe that these results are a reasonable indicator of the significance of inelastic effects. Because of the smallness of the effect, we have postponed further development of this aspect of the model.

Finally, a significant part of the effort on this program has been devoted to preparing a manuscript for publication that embodies a thorough analysis of the quantum kinetic theory of open systems. It is included here as Appendix D. The manuscript documents the studies (conducted largely under the previous program) that led us to the successful theory and presents a detailed analysis of this theory that considers both the fundamental continuum theory and the discretized model that must be employed for practical computations. This manuscript will be further expanded in a few areas before it is submitted for publication. We have not yet identified the proper forum for this work, as it is somewhat more tutorial and detailed than is customary for the standard research journals.

III. EXPERIMENT

The first report of laterally created discrete electronic states, work that was initiated under Contract No. N00014-84-C-0125 ("Research on GaAs Quantum-Coupled Structures That Can Be Used as Electron Devices"), was presented by us in an invited talk at the 1987 International Conference on Superlattices, Microstructures, and Microdevices. A preprint of this conference proceeding is attached as Appendix E. Subsequently, we developed under the current contract a better understanding of the potential(s) that confine discrete states in laterally defined resonant tunneling structures.

Because of the Fermi level pinning of the exposed GaAs surface, there is a narrow design window to observe discrete states; the physical size must be small enough to produce splittings greater than kT, but large enough so that pinch-off of the column does not occur. In fact, in the limit that the radius of the column equals the depletion width, the potential is perfectly parabolic and the column is just pinched off. Structures that exhibit the phenomena have a conduction path core (i.e., a difference between the radius of the column and the depletion width) typically less than 100 Å. Thus, the confining potential is essentially parabolic, with the eigenspectrum reflecting this parabolicity. The observation of the phenomena and the analysis was published in Physical Review Letters and is included here as Appendix F.

We have verified that the laterally confined discrete states arise from the quantum dot, not the contact, by fabricating structures similar to the quantum dot structures, except that the epitaxial structure only contains a single barrier; i.e., no quantum dot. No evidence of any fine structure from the quantized contact was observed in any of the numerous devices measured.

In both the single- and double-barrier structures, a previously reported "switching" phenomenon (impedance switching between two discrete values) due to single-electron trapping in these structures is often evident. We have shown that this phenomenon, sometimes called "telegraph noise," is the result of electrons trapped and emitted from impurity/defect states in the region of the double barriers. The change in impedance of the structure has been shown

to be caused by a modification of the potential in the double barrier region. This work has been submitted for publication and is included here as Appendix G. We also have evidence that this phenomenon is a result of the emptying and filling of clusters of interacting localized states in the double-barrier structure. At low temperatures we observe sharp two-level switching; at higher temperature the impedance takes on intermediate values, with the extrema still present. In addition, the distribution of switching times for both the "up" (high resistance) and "down" (low resistance") states follows a Lorentzian distribution, with the Lorentzian tail extending to long trapping times. This result is consistent with previous work that implicated clusters of traps.

An ongoing project is the creation of laterally confined states without the dominating depletion layers observed in the work referenced above. An alternative to the process described above is in situ etching and subsequent overgrowth of a heterojunction interface. We have initiated work on "thermal etching" (also called SUBLIME) for the creation of lateral resonant tunneling structures. While the technique has yet to show the submicrometer dimensions needed for this application, an even larger obstacle is the effect of the sustained elevated temperature on tunneling structures. We performed a study of the transport through double barrier structures that have been subjected to temperatures (while capped, to simulate MBE arsenic-stabilized surfaces) of up to 775°C (including 600°C, 675°C, 725°C, and 750°C) for up to eight hours. There is no appreciable change in the negative differential resistance (NDR) of these samples. Thus, the SUBLIME technique is viable for lateral resonant tunneling structures.

Finally, in modeling the above-referenced quantum dot structures, we have become concerned with the effect of contacts on understanding the spectroscopy of tunneling devices. We reverted to understanding the "simplest" case, a resonant tunneling diode (unconfined), to understand such effects as phonon scattering or relaxation on the position of the peak voltage. To test this we have grown, fabricated, measured, and modeled a series of highly characterized RTDs. The rather suprising and important results found are summarized as follows:

(a) In general, the model could not fit the voltage peak positions using the nominal values of the parameters (barrier thickness, QW thickness, etc.) for the characterized structures. Additionally, the variation of I-V asymmetry could not be explained.

(b) The positions, however, could be fit precisely if the parameters were varied within the measured error bars of the characterization. For example, the asymmetry (of the resonant peak positions) of the structures can be explained solely by different thicknesses of the top and bottom barriers.

The models we have developed are sufficiently well understood and precise, and other characterization techniques sufficiently imprecise, that these tunneling measurements and modeling results can serve as a diagnostic of the structure. These results, attached as Appendix H, were presented at the 15th International Symposium on GaAs and Related Compounds and will be published in Applied Physics Letters.

IV. CONCLUSIONS AND RECOMMENDATIONS

The results obtained in this contract indicate that quantum-coupled semiconductor structures display a rich variety of phenomena that can be exploited in a revolutionary post-VLSI IC technology. Such a technology will require interfaces between the macroscopic external world and the nanometer-scale devices. These interfaces and the role of contacts pose fundamental theoretical questions that are only partially answered. Significant fundamental work remains in developing realistic models of open quantum devices, nanofabrication, and exploration of electronic states in these devices. A number of promising avenues, such as laterally defined tunneling nanostructures, have yet to be explored.

Appendix A

"Quantum Transport Theory of Resonant-Tunneling Devices"

[To be published in Proceedings of the NATO Advanced Research Workshop on Band Structure Engineering in Semiconductor Microstructures (held at Il Ciocco, Italy, April 1988)]

QUANTUM TRANSPORT THEORY OF RESONANT TUNNELING DEVICES

William R. Frensley

Central Research Laboratories
Texas Instruments Incorporated
Dallas, Texas 75265

INTRODUCTION

The ability to fabricate semiconductor heterostructures on the scale of a few atomic layers has led to the development of devices which exploit the quantum-mechanical wave properties of electrons in their operation. The quantum device which has recieved the most attention recently is the quantum-well resonant-tunneling diode (RTD).[1,2] This device shows a negative-resistance characteristic which is quantum-mechanical in origin, and is potentially a very fast device. Most of the theoretical work on this device has employed the formal theory of scattering, focusing on the behavior of pure quantum states which are asymptotically plane waves. While this approach should adequately describe the device under stationary conditions, it is poorly equipped to treat any sort of time-varying behavior. The reason for this is that the behavior of the RTD, and indeed any electronic device, is manifestly time-irreversible, and a proper notion of irreversibility cannot be introduced into pure-state quantum mechanics. A pure quantum state cannot evolve time-irreversibly. Models which attempt to introduce such behavior inevitably violate some fundamental physical law, usually the continuity equation. However, *transitions* between quantum states may proceed irreversibly if the system of interest interacts with an external system having a continuum of states. Such processes may be consistently described in terms of statistically mixed states, which are represented most simply by the single-particle density matrix.[3] A description of a many-particle system in terms of such a single-particle distribution is generally termed a kinetic theory.[4] The present paper describes such a theory of electron devices which incorporates quantum coherence effects (including tunneling).

QUANTUM KINETIC TRANSPORT THEORY

A satisfactory transport theory must adequately treat two fundamental aspects of electron devices. First, an electron device is necessarily an open system; it is useless unless connected to an electrical circuit and able to exchange electrons with that circuit. If one wishes to study the behavior of the device apart from that of the circuit, it is convenient to represent the effects of the external circuit by ideal electron reservoirs attached to the terminals of the device. Secondly, a device is also a time-irreversible system, as evidenced by the set of nonequilibrium steady states which constitute the $I(V)$ characteristic. An elegant and consistent kinetic model of a device can be obtained which

incorporates *both* openness and irreversibility into the model via boundary conditions applied to the Wigner distribution function.[6] The Wigner function is simply a mathematical transform of the density matrix $\rho(x,x')$ so that it is defined in the phase-space (x,k_x):

$$f(x,k) = \int_{-\infty}^{\infty} dy\, e^{-ikx}\, \rho\left(x+\tfrac{1}{2}y,\, x-\tfrac{1}{2}y\right). \tag{1}$$

The Wigner function is often invoked to derive the correspondence between quantum and classical statistical mechanics. In the present case it provides the means by which an essentially classical model of the coupling of an open system to external reservoirs can be introduced into a quantum calculation. To describe purely ballistic transport of electrons (that is, neglecting collisions within the device) the Liouville equation for the time evolution of the Wigner function can then be written

$$\frac{\partial f}{\partial t} = -\frac{\hbar k}{m}\frac{\partial f}{\partial x} - \frac{1}{\hbar}\int_{-\infty}^{\infty}\frac{dk'}{2\pi}\, U(x,k-k')\, f(x,k') \equiv \frac{L}{i\hbar} f \tag{2}$$

where L is the Liouville super-operator. The form of the Liouville equation is quite similar to that in the classical case, with the exception that the effect of the potential is now non-local. This is how quantum interference effects enter the present model. The kernel of the potential operator is given by

$$U(x,k) = 2\int_0^{\infty} dy \sin(ky)\left[V\left(x+\tfrac{1}{2}y\right) - V\left(x-\tfrac{1}{2}y\right)\right] \tag{3}$$

The open system boundary conditions can be obtained in a physically appealing way by assuming that the reservoirs to which the device is connected have properties analogous to those of a black body: the distribution of electrons emitted into the device from the reservoir is characterized by the thermal equilibrium distribution function of the reservoir, and all electrons impinging upon a reservoir from the device are absorbed by the reservoir without reflection. To implement this picture, we must be able to distinguish the sense of the velocity of an electron at the position of the boundary. Thus the Wigner function is the natural representation for an open system, because it involves both the postion and the velocity. Let the interface between the device and the left-hand reservoir occur at $x=0$, and the interface between the device and the right-hand reservoir occur at $x=l$. Then we may write the open-system boundary conditions as

$$\begin{aligned} f(0,k)\big|_{k>0} &= F(k,\mu_l,T_l) \\ f(l,k)\big|_{k<0} &= F(k,\mu_r,T_r) \end{aligned} \tag{4}$$

where F is the Fermi distribution function (integrated over the transverse momenta), $\mu_{l,r}$ are the Fermi levels, and $T_{l,r}$ are the temperatures of the respective contacts. Note that these boundary conditions are in themselves time-irreversible, because under time-reversal they would map into a specification of the distribution of outgoing particles. While the Liouville equation (2) contains only the ballistic transport of electrons within the system, the irreversibility due to the coupling to the contacts is both necessary and sufficient to obtain a meaningful description of a device. Of course it will eventually be desirable to include those irreversible processes attributable to random scattering events, and a first approximation to such processes is described below.

The boundary conditions (4) are inhomogeneous. It is readily shown that the Liouville operator (2), subject to boundary conditions of the form (4) is non-singular, and its eigenvalues are confined to the lower half of the complex plane, corresponding to stable solutions.[7] Because L is non-singular, any choice of the boundary distribution F leads to a well-posed problelm. To obtain quantitative

results, the Wigner function is evaluated within a discrete (finite-difference and finite-sum) approximation.[5]

STEADY-STATE BEHAVIOR

The steady-state Wigner function is obtained by numerically solving the Liouville equation for the condition $\partial f/\partial t = 0$. The $I(V)$ characteristic for the RTD was obtained by calculating the steady-state Wigner function for each of a large set of bias voltages, and the current density was evaluated by averaging over the Wigner function. The results of such a calculation are shown in Fig. 1, along with the $I(V)$ curve obtained from a more conventional scattering-theory calculation for comparison. The agreement between these calculations is quite good in the vicinity of the peak tunneling current, and is somewhat poorer in the vicinity of the valley.

This agreement between the transport and scattering theories is considerably better than what was reported earlier.[5,7] The scattering calculations shown in the earlier work were in error, because the wrong velocity was used to evaluate the current density contribution from each state. The incorrect formula, which has been widely quoted,[8] involves the velocity of the electron on the incoming side of the barrier, so that this velocity cancels the "density of states" factor. The correct formula, as pointed out by Coon and Liu,[9] involves the velocity of the electron on the *outgoing* side of the barrier, which is not the same as the incoming velocity when there is a nonzero bias voltage. Correcting this error brings the scattering calculation into much better agreement with the predictions of the present quantum transport theory. Recent work by Mains and Haddad[10] indicates that modifications to the method of evaluating the potential operator (3) can have the effect of reducing the magnitude of the valley current. Such modifications should improve the agreement between the transport and scattering theories in this region of the $I(V)$ curve. These modifications have not yet been incorporated into the present calculations.

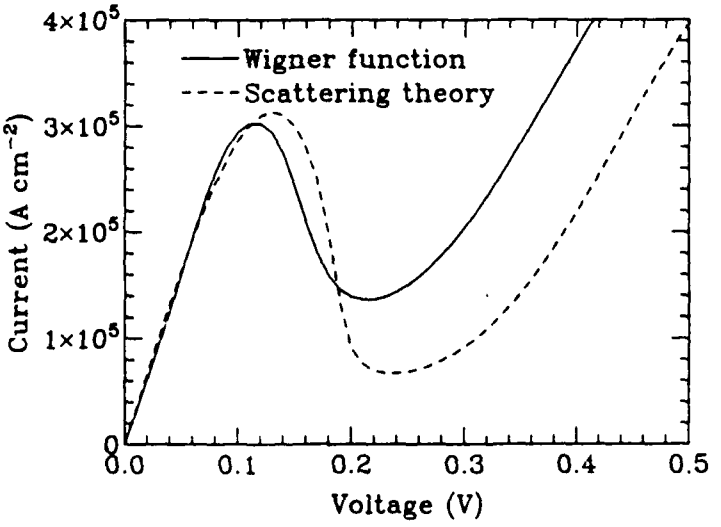


Fig. 1. Current density *vs.* voltage for a resonant-tunneling diode consisting of 2.8 nm layers of $Al_{0.3}Ga_{0.7}As$ bounding a 4.5 nm GaAs well, at a temperature of 300 K. The current derived from a calculation of the Wigner function (solid line) is compared to that derived from a more conventional scattering calculation (dashed line).

The device structure assumed in the present calculations consists of a 4.5 nm wide quantum well of GaAs bounded by identical 2.8 nm wide barrier layers of $Al_{0.3}Ga_{0.7}As$. The conduction-band discontinuity was taken to be 0.60 of the total bandgap discontinuity. 17.5 nm of the GaAs electrode layer was included in the simulation domain on each side of the device. Because Hartree self-consistency was not incorporated into the present calculations, the applied bias voltages were assumed to be dropped uniformly across the well and barriers. The electron density assumed in the boundary reservoirs was 2×10^{18} cm^{-3}. All calculations were performed at a temperature of 300 K.

TRANSIENT RESPONSE

The time-dependent response of the RTD to changes in the applied voltage is readily evaluated by integrating the Liouville equation (2), using the numerical procedures described in Ref. 5. The results of such a calculation are shown in Fig. 2. Since the negative-resistance characteristic is the interesting feature of this device, the transient response calculation was performed for a switching event across this region of the $I(V)$ curve. Figure 2 shows the current density in the device as a function of position and time for an event in which the initial bias of 0.11 V (corresponding to the peak in the current) was suddenly switched to 0.22 V (corresponding to the bottom of the valley) at $t=0$. More specifically, the steady-state Wigner function for a bias of 0.11 V was used as an initial value, and the time evolution under the Liouville operator for 0.22 V bias was evaluated. The response of the current is complex, as might be expected, but shows

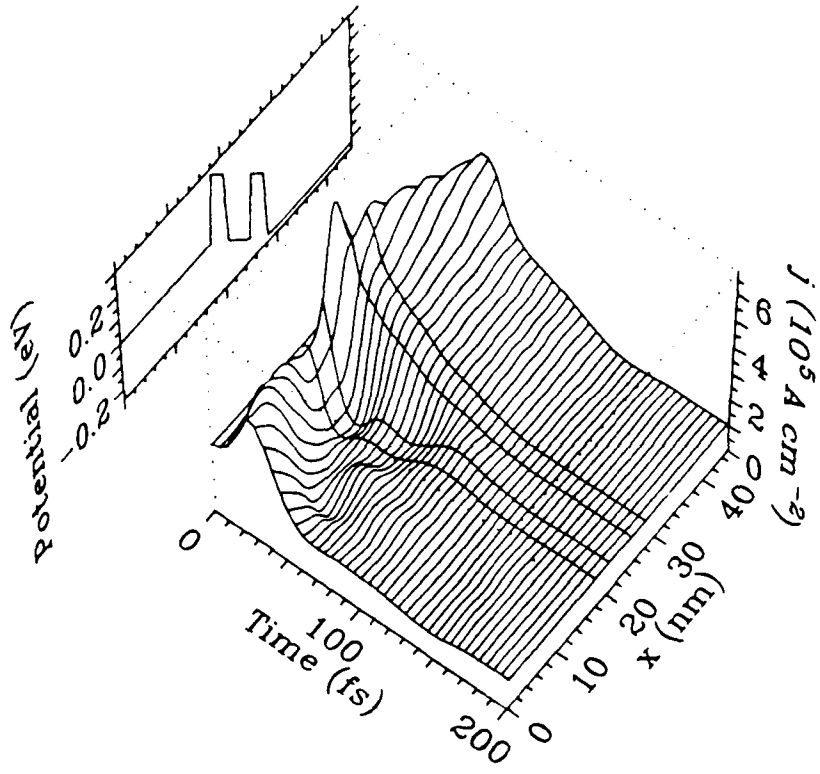


Fig. 2. Transient response of the resonant-tunneling diode of Fig. 1. Current density is plotted as a function of time and position within the device. The potential profile illustrates the device structure. At $t=0$, the voltage was suddenly switched from 0.11V (corresponding to the peak current) to 0.22V (corresponding to the valley current). After an initial peak, the current density approaches the lower steady-state value in 100-200 fs.

some features that are readily interpreted. The current density initially increases throughout the structure, so that the device displays a positive resistance over a short time. The destructive interference which underlies the negative resistance takes some tens of femtoseconds to manifest itself. The current has settled quite near to its steady-state value after 200 fs. Of course the response of real devices will be limited by the time required to charge the device capacitance through the parasitic series resistance of the contacts. Such effects were deliberately omitted from the present model in order to observe the intrinsic response of the tunneling process itself.

SMALL-SIGNAL RESPONSE

In order to obtain the small-signal ac response,[11] we assume that a small ac signal of amplitude v is superimposed upon the dc bias V. For a fixed V, the conduction current density j through the intrinsic device can be expanded in a power series in v, and to second order it is given by:

$$j(t) = j_0(V) + \tfrac{1}{2}(yve^{i\omega t} + cc) + \tfrac{1}{2}a_{rect}v^2 + \tfrac{1}{4}(a_{2\omega}v^2e^{2i\omega t} + cc) + \ldots , \tag{5}$$

where cc denotes the complex conjugate. Here ω is the angular frequency, j_0 is the dc current density, and y is the linear admittance (which equals dj_0/dV at $\omega = 0$). The nonlinear coefficients a_{rect} and $a_{2\omega}$ describe rectification and second-harmonic generation, respectively, and both are equal to d^2j_0/dV^2 at $\omega = 0$.

To obtain the small-signal ac response, we apply a simple form of perturbation theory to equation (2). The Liouville operator can be written as:

$$L = L_0 + \tfrac{1}{2}\lambda(L_\omega e^{i\omega t} + cc) . \tag{6}$$

The dc part L_0 includes the kinetic energy term and the dc potential. The ac part L_ω includes only the effect of the time-varying potential and thus is proportional to v. λ is a perturbation parameter introduced solely to keep track of the order of the perturbation, which will ultimately be set equal to unity. The Wigner function f can be expanded in a perturbation series, which to second order is given by

$$f = f_0 + \tfrac{1}{2}\lambda(f_\omega e^{i\omega t} + cc) + \lambda^2 f_{rect} + \tfrac{1}{2}\lambda^2(f_{2\omega}e^{2i\omega t} + cc) + \ldots . \tag{7}$$

Here f_0 is the dc part of the Wigner function, f_ω contains the linear ac response, and again f_{rect} and $f_{2\omega}$ describe rectification and second-harmonic generation, respectively. The perturbation equations are obtained by inserting (6) and (7) into (2) and collecting terms of equal frequency and equal order in λ. The resulting equations are:

$$L_0 f_0 = 0 . \tag{8a}$$

$$f_\omega = -\frac{1}{L_0 + \hbar\omega} L_\omega f_0 . \tag{8b}$$

$$f_{rect} = \frac{1}{2L_0} Re\left(L_\omega^* \frac{1}{L_0 + \hbar\omega} L_\omega f_0 \right) . \tag{8c}$$

$$f_{2\omega} = \tfrac{1}{2} \frac{1}{L_0 + 2\hbar\omega} L_\omega \frac{1}{L_0 + \hbar\omega} L_\omega f_0 \tag{8d}$$

These equations resemble those of the conventional perturbation series, but differ in detail primarily because quantum-mechanical convention for the time-dependence (e^{-iLt}) of f has been mixed with the electrical engineering convention ($e^{i\omega t}$) for the time-dependence of the applied signal. In particular, that is

the reason the "+" sign appears in the denominators. The resolvent expressions in (8b-d) are readily evaluated within the discretization approximation by ordinary matrix operations.

The contribution of a component f_i of the Wigner distribution to the terminal current density is obtained by averaging the current operator over the momentum, and over the active region of the device in accordance with the Ramo-Shockley theorem:[12,13]

$$\mathcal{J}[f_i] = \frac{1}{x_r - x_l} \int_{x_l}^{x_r} dx \int_{-\infty}^{\infty} \frac{dk}{2\pi} \frac{\hbar k}{m^*} f_i(x,k) \tag{9}$$

The coefficients in (4) are thus given by:

$$j_0 = \mathcal{J}[f_0] \tag{10a}$$

$$y = \mathcal{J}[f_\omega]/v \tag{10b}$$

$$a_{rect} = \tfrac{1}{2}\mathcal{J}[f_{rect}]/v^2 \tag{10c}$$

$$a_{2\omega} = \tfrac{1}{2}\mathcal{J}[f_{2\omega}]/v^2 \tag{10d}$$

The small-signal response was evaluated for an assumed structure which was the same as that described above, except that the doping in the contact layers was taken to be 2×10^{17} cm^{-3}. This structure is similar to the sample number 2 of Sollner *et al.*[14] The linear admittance was evaluated from (8b) and (10b), for a bias voltage near the center of the negative resistance region. The resulting admittance as a function of frequency is shown in Fig. 3. The real conductance is negative at lower frequencies, as expected. The negative conductance "rolls off" in the THz region and goes positive at about 6 THz. The imaginary part of the electronic admittance is negative and proportional to ω at lower frequencies, and thus resembles an inductance. This is due to the phase shift resulting from the electrons' inertia.[15] The rather complex behavior of the

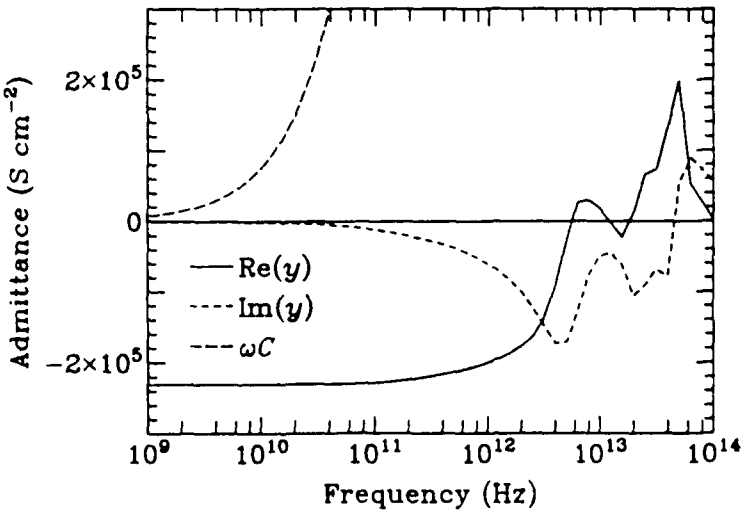


Fig. 3. Electron admittance as a function of frequency. The electron conductance is Re(y) and the electron susceptance, due to inertial effects, is Im(y). The negative conductance at lower frequencies is apparent. The susceptance due to the parasitic capacitance ωC is shown to provide a measure of the effect of the parasitic elements.

electronic admittance above 10 THz reflects other resonant processes in the system. In this frequency range the current response is quite nonlocal.[16]

An estimate of the susceptance of the parasitic capacitance of the RTD is also plotted in Fig. 3, for comaprison. The effects of this capacitance will become dominant when the magnitude of its admittance exceeds that of the tunneling current, which occurs, for the present model, somewhat below 100 GHz. This is the practical limit for the observation of a linear negative conductance.

Some nonlinear effects are observable to much higher frequencies than the linear effects, however. To examine the behavior of such processes, the nonlinear coefficients were evaluated from (8c,d) and (10c,d) for that dc voltage at the resonant peak of the $j(V)$ curve. The modulus of a_{rect} and of $a_{2\omega}$ are plotted in Fig. 4 as functions of frequency. The interesting point is that the calculations predict an enhancement in the coefficient for rectification between 1 and 8 THz. This agrees with the observations of Sollner *et al.*[2] of rectification at 2.5 THz in their experi-mental devices. The quantity a_{rect} is the same as that which is denoted I'' in Ref. 2.

EFFECTS OF PHONON SCATTERING

The effects of the electron-phonon interaction may be easily incorporated into the present transport theory by adding an appropriate collision superoperator C to the Liouville equation:

$$\frac{\partial f}{\partial t} = \frac{L}{i\hbar} f + Cf \tag{11}$$

The existing numerical machinery can handle such a term so long as C can be treated as local in space and time. The obvious first step toward obtaining such an operator is to employ the classical Boltzmann equation form:

$$[Cf](x,k,t) = \int dk' \, [W_{k\,k'} f(x,k',t) - W_{k'\,k} f(x,k,t)], \tag{12}$$

where $W_{k\,k'}$ is the transition rate from k' to k, *etc.* The work of Levinson[17] and

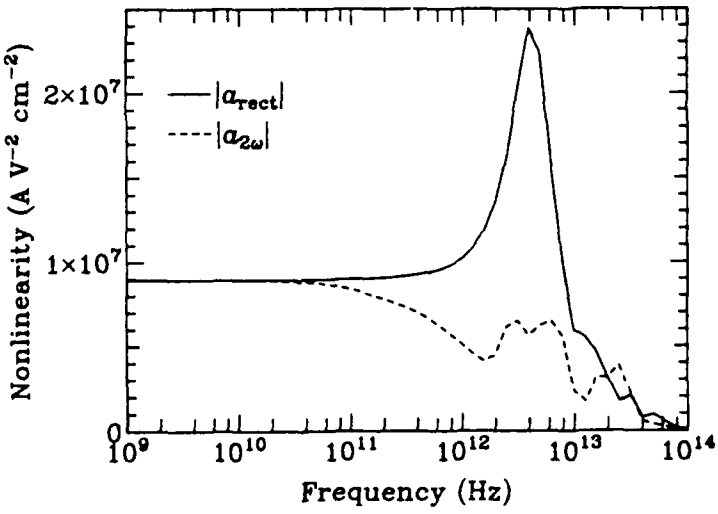


Fig. 4. The nonlinear response coefficients as functions of frequency. Rectification is described by $a_{\rm rect}$ and second-harmonic generation is described by $a_{2\omega}$. The persistence of the rectification effect to terahertz frequencies is in agreement with the experimental results of Ref. 2.

that of Lin and Chu[18] suggests that this is an appropriate approximation in the semi-classical case (that is, when the Fermi golden rule may be used).

Because the RTD is in reality a three-dimensional system, the integral in (12) must be three-dimensional. However, the model is one dimensional, so the operator in (12) must be projected onto the one-dimensional subspace by integrating over the transverse wavevectors $\mathbf{k}_\perp$ and $\mathbf{k}'_\perp$. Invoking once more the assumption that the distributions with respect to the transverse momenta are Boltzmann, the expression for the transition rates projected into one dimension is:

$$W_{k\,k'} = \frac{2\pi}{\hbar}\frac{V}{(2\pi)^3}\int d^2\mathbf{k}_\perp \int d^2\mathbf{k}'_\perp |\langle \mathbf{k}|H'|\mathbf{k}'\rangle|^2 \delta(E_\mathbf{k} - E_{\mathbf{k}'} \mp \hbar\omega)\,\frac{2\pi\beta\hbar^2}{m}\,exp\left|-\frac{\beta\hbar^2\mathbf{k}'^2_\perp}{2m}\right| \quad (13)$$

Here H' is the Hamiltonian which describes the particular electron-phonon interaction. The numerical collision operators obtained from (12) and (13) were checked for consistency with the requirements of detailed balance by applying the operator to an equilibrium distribution function and verifying that the result was zero.

In the present calculations the deformation potential interaction was included for scattering with acoustic phonons and the Fröhlich interaction was included for scattering with longitudinal optical (LO) phonons.[19] The effects of these phonon scattering mechanisms on the $I(V)$ curve of the RTD are shown in Fig. 5. Acoustic phonon scattering has a nearly negligible effect on the $I(V)$ curve. The effect of LO phonon scattering is rather more pronounced, primarily in the reduction of the peak current. When both acoustic and LO phonons are included in the calculation, the resulting $I(V)$ curve is indistinguishable from that obtained with LO scattering only.

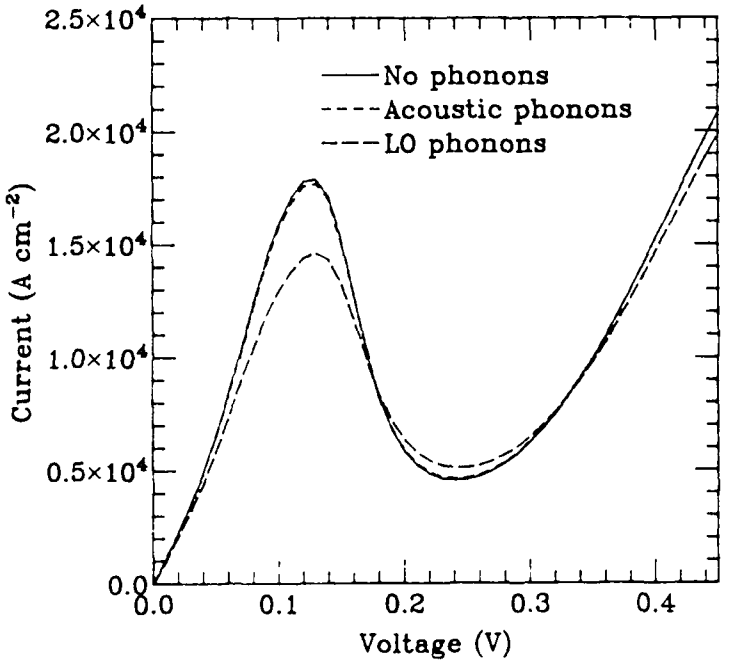


Fig. 5. Effect of semi-classical phonon-scattering operators on the $I(V)$ characteristic of a resonant-tunneling diode.

CONCLUSIONS

The present quantum kinetic transport theory has proven to be exceptionally effective in modeling the behavior of resonant-tunneling devices. This success may be attributed to several key elements: First, it incorporates a simple and explicit notion of irreversibility through the coupling of the device to its contacts. This permits a simple treatment of irreversible phenomena such as the transient response, which are beyond the scope of theories which do not explicitly include irreversibility. The second element is that it corresponds as closely as possible to a classical model, departing only as required to include quantum interference effects [which enter via the nonlocal potential (3)]. This permits a direct computation of classically measurable quantities, such as the small-signal admittance. Finally, in contrast to the more sophisticated techniques of many-body theory, it suppresses enough of the complexities of the system so as to remain computationally tractable.

ACKNOWLEDGEMENT

This work was supported by the U.S. Office of Naval Research.

REFERENCES

1. L. L. Chang, L. Esaki and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).
2. T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker and D. D. Peck, Appl. Phys. Lett. **43**, 588 (1983).
3. U. Fano, Rev. Mod. Phys. **29**, 74 (1957).
4. Kubo, R., M. Toda, and N. Hashitsume, *Statistical Physics II. Nonequilibrium Statistical Mechanics.* (Springer-Verlag, Berlin, 1985).
5. W.R. Frensley, Phys. Rev. **B36**, 1570 (1987).
6. E. Wigner, Phys. Rev. **40**, 749 (1932).
7. W.R. Frensley, Phys. Rev. Lett. **57**, 2853 (1986).
8. R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).
9. D.D. Coon and H.C. Liu, Appl. Phys. Lett. **47**, 172 (1985).
10. R.K. Mains and G.I. Haddad, "Numerical Considerations in the Wigner Function Modeling of Resonant-Tunneling Diodes," to be published.
11. W.R. Frensley, Appl. Phys. Lett. **51**, 448 (1987).
12. S. Ramo, *Proc. IRE* **27**, 584 (1939).
13. W. Shockley, *J. Appl. Phys.* **9**, 635 (1938).
14. T.C.L.G. Sollner, E.R. Brown, W.D. Goodhue, and H.Q. Le, *Appl.Phys. Lett.* **50**, 332 (1987).
15. K.S. Champlin, D.B. Armstrong, and P.D. Gunderson, *Proc. IEEE* **52**, 677 (1964).
16. W.R. Frensley, to be published in Superlattices and Microstructures.
17. I.B. Levinson, "Translational invariance in uniform fields and the equation of motion for the density matrix in the Wigner representation," Soviet Physics JETP **30**, 362-7 (1970).
18. J. Lin and L.C. Chiu, "Quantum theory of electron transport in the Wigner formalism," J. Appl. Phys. **57**, 1373-6 (1985).
19. E.M. Conwell, *High Field Transport in Semiconductors*, (Academic Press, New York, 1967) ch. 5.

Appendix B

"Quantum Transport Calculation of the Frequency Response of Resonant-Tunneling Heterostructure Devices"

[Published in Superlattices and Microstructures 4, 4/5, 497 (1988)]

QUANTUM TRANSPORT CALCULATION OF THE FREQUENCY RESPONSE OF RESONANT-TUNNELING HETEROSTRUCTURE DEVICES

William R. Frensley
Central Research Laboratories
Texas Instruments Incorporated
P.O. Box 655936, MS 154
Dallas Texas 75265

(Received 17 August 1987)

The frequency response of the quantum-well resonant-tunneling diode is calculated using quantum transport theory. The state of the device is represented by the single-particle Wigner distribution function. The Wigner function is obtained by numerical solution of the Liouville equation, subject to inhomogeneous boundary conditions that represent the ohmic contacts to the device and that introduce dissipation into the model. The small-signal ac response is calculated by a perturbation expansion about the non-equilibrium steady state. The calculations indicate that both the negative conductance and nonlinear rectification persist up to the mid terahertz region, for the design studied. This is compared to the lifetime of the resonant state inferred from the width of the scattering resonance.

Introduction

The quantum-well resonant-tunneling diode (RTD)[1,2] is the simplest semiconductor heterostructure that displays interesting device properties due to quantum coherence effects. It is thus an ideal prototype system for which to develop techniques for the analysis of quantum devices. A form of quantum transport theory has been developed that is adapted to the study of quantum devices because it provides a means of treating the electrical contacts to the device.[3,4] Recent interest in the RTD can be attributed to the work of Sollner *et al.*,[2] who demonstrated nonlinear electrical response in these devices at frequencies up to 2.5 THz. The existence of these results provides a motivation for the development of theoretical techniques to evaluate the small-signal ac response of a tunneling device. The present work demonstrates that such calculations may be readily performed by applying the techniques developed in Refs. 3 and 4.

Transport Model

The physical model of the resonant-tunneling diode is summarized in Fig. 1. The device is considered to be a finite region of semiconductor, characterized by a potential V that includes the effects of applied voltages and of heterojunction band offsets. The boundaries of the device are taken to be interfaces to particle reservoirs, by which the terminals

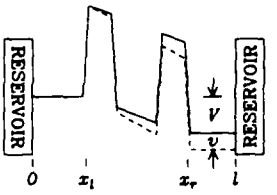


Figure 1. Physical model of the RTD, showing the device potential. A dc bias V and a small ac signal of amplitude v are applied to the intrinsic device.

of the device are modeled. The internal state of the device is represented by the Wigner distribution function,[5] which is the quantum analog of the classical distribution function that appears in the Boltzmann equation.

Regarding the contacts as particle reservoirs gives a well-defined model of the open-system nature of the device. The interaction between the reservoir and the device is simply described: Electrons in the device that impinge upon the reservoir pass into the reservoir without reflection and the distribution of electrons entering the device from the reservoir is given by the equilibrium distribution of the reservoir. These boundary conditions are a crucial aspect of the model, because they permit the existence of steady-state solutions under applied bias and they lead to a stable approach

0749-6036/88/040497+05 $02.00/0

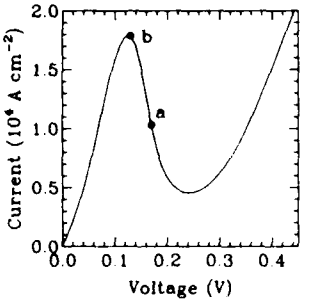


Figure 2. The current-voltage curve derived from the steady-state Wigner function calculation. The linear response shown in Fig. 4 was evaluated at the dc bias shown as point "a," and the nonlinear response of Fig. 6 was evaluated at point "b."

to those steady-state solutions after the bias voltage is changed.

The time evolution of the Wigner function is described by the Liouville equation

$$\frac{\partial f}{\partial t} = \frac{L}{i\hbar} f = -\frac{\hbar k}{m}\frac{\partial f}{\partial x} - \frac{1}{\hbar}\int_{-\infty}^{\infty}\frac{dk'}{2\pi} U(x, k-k') f(x, k') \qquad (1)$$

where L is the Liouville super-operator and the kernel of the potential operator U is given by

$$U(x,k) = 2\int_0^{\infty} dy \sin(ky)[V(x+\tfrac{1}{2}y) - V(x-\tfrac{1}{2}y)] \qquad (2)$$

The open-system boundary conditions are

$$\begin{aligned} f(0,k) &= F(\mu_l, T_l) \quad k>0 \\ f(l,k) &= F(\mu_r, T_r) \quad k<0 \end{aligned} \qquad (3)$$

where F is the Fermi distribution function (integrated over the transverse momenta), $\mu_{l,r}$ are the Fermi levels, and $T_{l,r}$ are the temperatures of the respective contacts.

Equation (1) is discretized on a uniform mesh in the phase space (x,k). The boundary conditions lead to a natural discretization of the gradient term with a left-hand difference for $k>0$ and a right-hand difference for $k<0$. This is an "upwind" difference and is the means by which the boundary conditions stabilize the solutions of the Liouville equation. The Liouville equation (1) is readily solved for the steady-state condition $\partial f/\partial t = 0$, subject to the inhomogeneous boundary conditions (3). This is done for a set of bias voltages and the current is evaluated from the Wigner function to obtain an $I(V)$ curve as illustrated in Fig. 2.

Small Signal Response Theory

To obtain the small-signal ac response, we assume that a small ac signal of amplitude v is superimposed upon the dc bias V. For the purpose of obtaining the electrostatic potential, the contact layers on either side of the quantum-well barriers are assumed to be ideally metallic (i.e., the accumulation and depletion layers are taken to be of infinitesimal width). The equivalent circuit of the device is shown in Fig. 3.[6-8] The series resistance R_S is due to the combined effects of all contacting layers, semiconducting and metallic, and as such is a quantity that depends purely on the device design and fabrication technology. The capacitance C is due to the depletion of electrons in the vicinity of the quantum well structure. The current source j responds to the voltage applied across it, and represents the electronic response of the intrinsic device. For a fixed dc bias voltage V, the current density j can be expanded in a power series in v, and to second order it is given by:

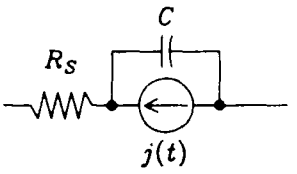


Figure 3. Equivalent circuit of the RTD. Electron conduction through the intrinsic device is represented by a current source, whose specification is the purpose of this paper. A parallel displacement current flows through the parasitic capacitance C. The series resistance R_S represents the effects of the contacts.

$$j(t) = J_0(V) + \tfrac{1}{2}(\lambda v e^{i\omega t} + cc) + \tfrac{1}{4}(a_{rect} v^2 + \tfrac{1}{2}(a_{2\omega} v^2 e^{2i\omega t} + cc)) \qquad (4)$$

where cc denotes the complex conjugate. Here ω is the angular frequency, J_0 is the dc current density, and y is the linear admittance (which equals dJ_0/dV at $\omega = 0$). The nonlinear coefficients a_{rect} and $a_{2\omega}$ describe rectification and second-harmonic generation, respectively, and both are equal to d^2J_0/dV^2 at $\omega = 0$.

To obtain the small-signal ac response, we apply a simple form of perturbation theory to equation (1). The Liouville operator can be written as:

$$L = L_0 + \tfrac{1}{2}\lambda(L_\omega e^{i\omega t} + cc) \qquad (5)$$

The dc part L_0 includes the kinetic energy term and the dc potential as shown in Fig. 1. The ac part L_ω includes only the effect of the time-varying potential and thus is proportional to v. λ is a perturbation parameter introduced solely to keep track of the order of the perturbation, which will ultimately be set equal to unity. The Wigner function f can be expanded in a perturbation series, which to second order is given by

$$f = f_0 + \tfrac{1}{2}\lambda(f_\omega e^{i\omega t} + cc) + \lambda^2 f_{rect} + \tfrac{1}{2}\lambda^2(f_{2\omega} e^{2i\omega t} + cc) + \ldots \quad (6)$$

Here f_0 is the dc part of the Wigner function, f_ω contains the linear ac response, and again f_{rect} and $f_{2\omega}$ describe rectification and second-harmonic generation, respectively. The perturbation equations are obtained by inserting (5) and (6) into (1) and collecting terms of equal frequency and equal order in λ. The resulting equations are:

$$L_0 f_0 = 0 \quad (7a)$$

$$f_\omega = -\frac{1}{L_0 + \hbar\omega} L_\omega f_0 \quad (7b)$$

$$f_{rect} = \frac{1}{2L_0} Re\left(L_\omega^* \frac{1}{L_0 + \hbar\omega} L_\omega f_0\right) \quad (7c)$$

$$f_{2\omega} = \tfrac{1}{2} \frac{1}{L_0 + 2\hbar\omega} L_\omega \frac{1}{L_0 + \hbar\omega} L_\omega f_0 \quad (7d)$$

The resolvent expressions in (7b-d) are readily evaluated within the discretization approximation by ordinary matrix operations.

The contribution of a component f_i of the Wigner distribution to the terminal current density is obtained by averaging the current operator over the momentum, and over the active region of the device in accordance with the Ramo-Shockley theorem[9,10]:

$$J(f_i) = \frac{1}{x_r - x_l} \int_{x_l}^{x_r} dx \int_{-\pi}^{\pi} \frac{dk}{2\pi} \frac{\hbar k}{m^*} f_i(x,k) \quad (8)$$

The coefficients in (4) are thus given by:

$$j_0 = J(f_0) \quad (9a)$$

$$y = J(f_\omega)/v \quad (9b)$$

$$a_{rect} = \tfrac{1}{2} J(f_{rect})/v^2 \quad (9c)$$

$$a_{2\omega} = \tfrac{1}{2} J(f_{2\omega})/v^2 \quad (9d)$$

Predicted Small-Signal Response

The device structure assumed in the present calculations is similar to the sample number 2 of Sollner *et al.*[6] The model structure consists of a 4.5 nm wide quantum well of GaAs bounded by identical 2.8 nm barrier layers of $Al_{0.3}Ga_{0.7}As$. GaAs electrode layers 17.5 nm wide and doped at 2×10^{17} cm^{-3} were included in the simulation domain on each side of the device. All calculations were performed for a temperature of 300 K. The dc $j(V)$ curve was evaluated by solving (5a) as described in Ref. 4, and the result is shown in Fig. 2.

The linear admittance was evaluated from (7b) and (9b), for $V = 0.17$ V, which is near the center of the negative resistance region. The resulting admittance as a function of frequency is shown in Fig. 4. The real conductance is negative at lower frequencies, as expected. The negative conductance "rolls off" in the THz region and goes positive at about 6 THz. The imaginary part of the electronic admittance is negative and proportional to ω at lower frequencies, and thus resembles an inductance. This is due to the phase shift resulting from the electrons' inertia.[11] This inductance, however, is five orders of magnitude too small to explain the inductance measured by Gering *et al.*[7]

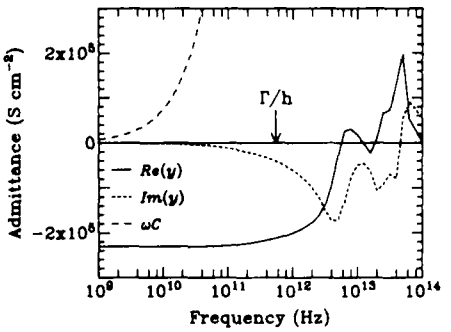


Figure 4. Electron admittance as a function of frequency. The electron conductance is Re(y) and the electron susceptance, due to inertial effects, is Im(y). The negative conductance at lower frequencies is apparent. The susceptance due to the parasitic capacitance ωC is shown to provide a measure of the effect of the parasitic elements.

To provide an understanding of the role of the parasitic elements of Fig. 3, the capacitive susceptance ωC is also plotted in Fig.4. If we represent the tunneling current by its conductance $g = Re(y)$ [and neglect $Im(y)$] then the resistance of the parallel conductance and capacitance is[7,8]

$$R(g\|C) = g^{-1}[1 + (\omega C/g)^2]^{-1} \quad (10)$$

In the frequency range of interest g is negative, leading to a negative resistance which rolls off when $|\omega C/g| \approx 1$. From Fig. 4 it is apparent that this will occur around 40 GHz, well below the cutoff frequency of the tunneling current itself. As shown in Ref. 3, this leads to the "circuit limit" on the maximum oscillation frequency, which is reached when the negative resistance from (10) can no longer cancel the series resistance R_S. The capacitance per unit area was estimated by $\varepsilon/(x_r - x_l)$, to be consistent with the assumed form of the potential. The true capacitance will be lower when the finite depletion layer width is included. Nevertheless, the present 40 GHz corner frequency is comparable to the observed f_{max} of experimental devices.[6]

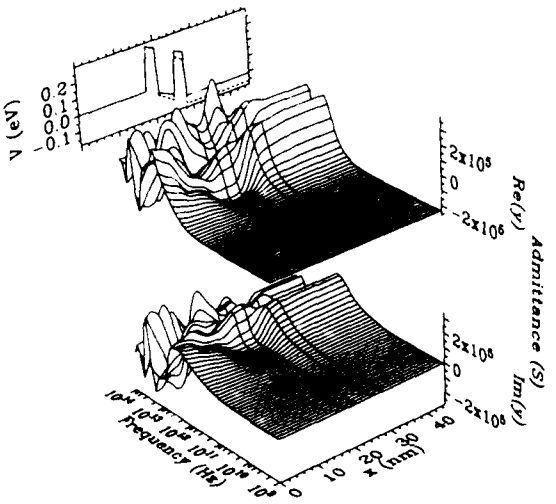


Figure 5. Linear component of the ac current density (divided by the ac voltage and thus expressed as an admittance) as a function of frequency and position. The ac electric field is again assumed to be uniform over the intrinsic device. The nonlocality of the current response at higher frequencies is evident.

The rather complex behavior of the electronic admittance above 10 THz reflects other resonant processes in the system. In this frequency range the current response is quite nonlocal. This is demonstrated in Fig. 5, where the same results as in Fig. 4 are now resolved with respect to the position x. The presence of complicated resonance phenomena above 10 THz is apparent. Also, the negative conductance appears to persist to somewhat higher frequencies in the vicinity of the left-hand barrier than in other parts of the structure. This perhaps reflects the process of filling and emptying the well by current through this barrier.

The nonlinear coefficients were evaluated from (7c,d) and (9c,d) for $V = 0.13$ V at the resonant peak of the $j(V)$ curve. The modulus of a_{rect} and of $a_{2\omega}$ are plotted in Fig. 6 as functions of frequency. The interesting point is that the calculations predict an enhancement in the coefficient for rectification between 1 and 8 THz. This agrees with the observations of Sollner *et al.*[2] of rectification at 2.5 THz in their experimental devices. The quantity a_{rect} is the same as that denoted I'' in Ref. 2.

Discussion

From the forms of Eqs. (7) it is apparent that the eigenvalues of L_0 produce poles in the frequency response, and the numerical calculations imply

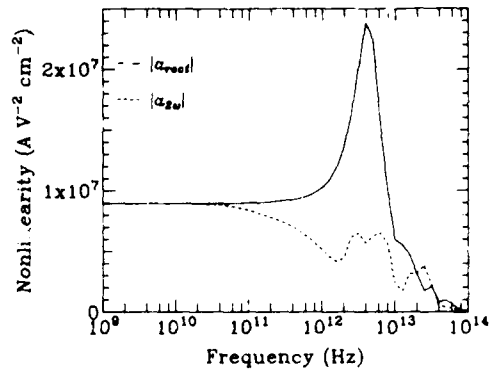


Figure 6. The nonlinear response coefficients as functions of frequency. Rectification is described by a_{rect} and second-harmonic generation is described by $a_{2\omega}$. The persistence of the rectification effect to terahertz frequencies is in agreement with the experimental results of Ref. 2.

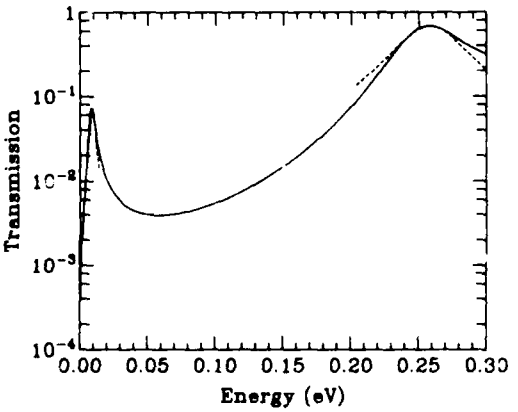


Figure 7. The transmission probability $|T|^2$ as a function of energy for the assumed device structure with a bias of 0.17 V. Breit-Wigner resonance forms (dashed lines) were fit to the resonant peaks to estimate the resonant state lifetime.

that the smallest eigenvalues correspond to submillimeter-wave frequencies. Time-domain calculations of the full transient response[3,4,12] also indicate that the smallest eigenvalues of $L_0/\hbar$ are of the order of 10^{13} s^{-1} for the present structure.

This may be compared to the frequently invoked[13-15] delay time estimate $\hbar/\Gamma$, where Γ is the width of the resonant peak in the transmission amplitude. Γ was determined for the present structure by a numerical calculation of the scattering transmission amplitude. The Schroedinger equation was discretized with respect to x on the same

spatial mesh as was used to discretize the Liouville equation. The wave function was then calculated recursively[16] for a range of incident energies. (The bias voltage was taken to be 0.17 V, as in the admittance calculation. The resulting transmission probability as a function of energy is shown in Figure 7. A Breit-Wigner[17] expression of the form

$$|T|^2 = |T_0|^2 \Gamma^2/[(E-E_0)^2+\Gamma^2] \qquad (11)$$

was fit to the resonant peak. The value of Γ was thus determined to be 2.27 meV, leading to an estimate of the cutoff frequency $\Gamma/h = 5.5\times10^{11}$ Hz. This estimate is plotted in Fig. 4 for comparison with the admittance calculation. Γ/h gives a frequency that is about an order of magnitude lower than the transport-theory result for the cutoff frequency. A more detailed investigation of the behavior of both the eigenvalues of L_0 and Γ as functions of barrier thickness will be required to clearly display the relationship between these quantities.

Acknowledgement—This work was supported in part by the Office of Naval Research and the Army Research Office.

References

1. L.L. Chang, L. Esaki and R. Tsu, *Applied Physics Letters* **24**, 593 (1974).
2. T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker and D. D. Peck, *Applied Physics Letters* **43**, 588 (1983)
3. W.R. Frensley, *Physical Review Letters* **57**, 2853 (1986).
4. W.R. Frensley, *Physical Review B* **36**, 1570 (1987).
5. E. Wigner, *Physical Review* **40**, 749-59 (1932).
6. T.C.L.G. Sollner, E.R. Brown, W.D. Goodhue, and H.Q. Le, *Applied Physics Letters* **50**, 332 (1987).
7. J.M. Gering, D.A. Crim, D.G. Morgan, P.D. Coleman, W. Kopp, and H. Morkoç, *Journal of Applied Physics* **61** 271 (1987).
8. B. Jogai, K.L. Wang, and K.W. Brown, *Applied Physics Letters* **48**, 1003 (1986).
9. S. Ramo, *Proceedings of the IRE* **27**, 584 (1939).
10. W. Shockley, *Journal of Applied Physics* **9**, 635 (1938).
11. K.S. Champlin, D.B. Armstrong, and P.D. Gunderson, *Proceedings of the IEEE* **52**, 677 (1964).
12. W.R. Frensley, *IEEE International Electron Device Meeting, Technical Digest* **1986**, p. 571.
13. B. Ricco and M. Ya. Azbel, *Physical Review B* **29**, 1970 (1984).
14. D.D. Coon and H.C. Liu, *Applied Physics Letters* **49**, 94 (1986).
15. P.J. Price, *Superlattices and Microstructures* **2**, 593 (1986).
16. J.P. Vigneron and Ph. Lambin, *Journal of Physics A: Mathematical and General*, **13**, 1135 (1980).
17. G. Breit and E. Wigner, *Physical Review* **49**, 519 (1936).

Appendix C

"Quantum Transport Modeling of Resonant-Tunneling Devices"

[Published in Solid-State Electronics 31, 3/4, 739 (1988)]

Solid-State Electronics Vol. 31, No. 3/4, pp. 739–742, 1988
Printed in Great Britain

0038-1101/88 $3.00 + 0.00
Pergamon Journals Ltd

QUANTUM TRANSPORT MODELING OF RESONANT-TUNNELING DEVICES

William R. Frensley

Texas Instruments Incorporated
P.O. Box 655936, M.S. 154
Dallas, Texas 75265

ABSTRACT

A form of quantum transport theory has been developed to model the resonant-tunneling diode and similar devices in which quantum interference effects play a significant role. The internal state of the device is represented by the Wigner distribution function, with boundary conditions which model the effects of the electrical contacts to the device. Inelastic scattering processes are approximated by a classical Boltzmann collision operator, and the effects of different scattering processes on the device characteristics are evaluated numerically.

KEYWORDS

Quantum transport; Wigner distribution function; Resonant tunneling; Electron-phonon interaction; Transient response; Nanoelectronics.

INTRODUCTION

The progress of semiconductor fabrication technology has permitted the fabrication of devices whose behavior is dominated by quantum-interference effects. The most widely studied example of a quantum size-effect device is the resonant tunneling diode (RTD) (Chang, Esaki, and Tsu, 1974; Sollner and colleagues, 1983). This device exhibits interesting properties in the form of a negative-resistance region of its characteristic curve which unambiguously shows the quantum-mechanical nature of electron transport through this structure. The central issue in the theory of such devices concerns the proper description of the dissipative processes which determine their behavior. Such processes can be grouped into two categories: interactions of conduction electrons with other kinds of particles in the crystal (such as phonons), and the exchange of those electrons with the elements of the external electrical circuit. Previous work (Frensley 1986, 1987a, 1987b) has demonstrated that a consistent model of a tunneling device may be obtained by invoking only the latter type of interaction. The present paper extends this model to include phonon scattering.

TRANSPORT MODEL

In the present model the device is considered to be a finite region of semiconductor, characterized by a potential that includes the effects of applied voltages and of heterojunction band offsets. The internal state of the device is represented by the single-particle Wigner distribution function $f(x,k,t)$ where x is the position and k is the momentum (Wigner, 1932). The Wigner function is assumed to obey a Markovian kinetic equation of the form

$$\frac{\partial f}{\partial t} = \frac{L}{i\hbar} f + Cf \tag{1}$$

where L is the Liouville super-operator which describes ballistic electron motion and C is a collision super-operator which describes random scattering. The boundary conditions on f describe the coupling of the device to electron reservoirs which model the electrical contacts to the device. The boundary conditions specify only the distribution of electrons *entering* the device, and thus introduce time-irreversibility into the model independently of the collision operator. The Wigner function f is calculated in a discrete approximation, which reduces the integro-differential equation (1) to a large system of linear algebraic equations which are solved numerically (Frensley, 1987a).

The collision super-operator C is assumed to be of the classical form:

$$[Cf](x,k,t) = \int dk' \{W_{k,k'} f(x,k',t) - W_{k',k} f(x,k,t)\}, \tag{2}$$

where $W_{k,k'}$ is the transition rate from k' to k, *etc.*

SIMPLE COLLISION MODELS

There are two readily available approximations for the collision operator C, both of which have the form (2). One is the well-known relaxation term:

$$[C_{relax} f](k) = (1/\tau)[f_0(k) \int dk' f(k') - f(k)], \tag{3}$$

where τ is a relaxation time, and f_0 is the normalized equilibrium distribution function. The other simple approximation to C is the dissipation operator derived from the Fokker-Planck or Kramers equation for Brownian motion (Caldeira and Leggett, 1983; Kubo, Toda, and Hashitsume, 1985):

$$[C_{FP} f](k) = -\frac{1}{\tau}\left[\frac{\partial}{\partial k}(kf) + \frac{m}{\beta\hbar^2}\frac{\partial^2 f}{\partial k^2}\right]. \tag{4}$$

This approximation assumes a very high rate of low momentum-transfer collisions.

These simple collision terms were tested in numerical calculations of the dc $I(V)$ curve of a resonant-tunneling diode, by solving (1) for steady state. A relaxation time τ of 100 fs was assumed for both the relaxation and Fokker-Planck models, corresponding to a mobility of 2600 $cm^2V^{-1}s^{-1}$ for GaAs. The resulting characteristic curves are shown in Fig. 1. The relaxation term (3) greatly reduces the peak-to-valley current ratio, both decreasing the peak current and increasing the valley current. The Fokker-Planck term (4), however, leads to an increase in the current at all voltages.

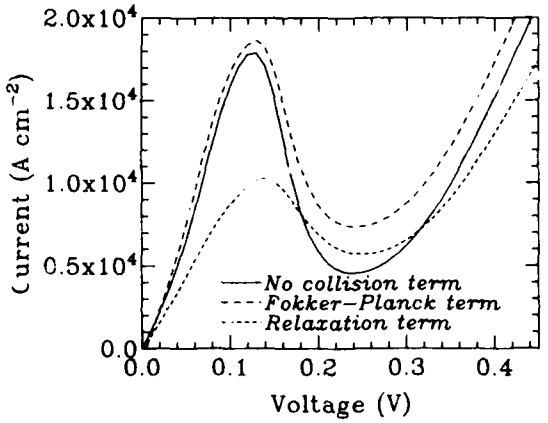


Fig. 1. Effect of simple collision terms on the $I(V)$ characteristic of a resonant-tunneling diode.

REALISTIC PHONON SCATTERING

The sensitivity of the $I(V)$ curve to the form of the collision operator demonstrated by these simple models implies that we need to use more realistic models of random scattering processes in the device. An obvious first step in this direction is to retain the classical Boltzmann form (2), and use the Fermi golden rule to calculate the transition rates W. The work of Levinson (1970) and of Lin and Chiu (1985) suggests that this is an appropriate approximation when the electron-phonon interaction can be treated semi-classically. Because only a single spatial dimension is resolved in the numerical model, we must make some assumptions about the dependence on the transverse components of the momentum $\mathbf{k}$. The obvious assumption is that the distribution is Maxwellian with respect to the transverse wavevector $\mathbf{k}'_\perp$. Then integrating the resulting distribution function with respect to $\mathbf{k}_\perp$, we obtain

$$W_{k'k} = \frac{2\pi}{\hbar}\frac{V}{(2\pi)^3}\int d^2\mathbf{k}_\perp \int d^2\mathbf{k}'_\perp |\langle \mathbf{k}|H'|\mathbf{k}'\rangle|^2 \delta(E_{\mathbf{k}} - E_{\mathbf{k}'} \mp \hbar\omega) \frac{2\pi\beta\hbar^2}{m} \exp\left[\frac{\beta\hbar^2 \mathbf{k}'^2_\perp}{2m}\right] \tag{5}$$

Here H' is the Hamiltonian which describes the particular electron-phonon interaction. The numerical collision operators obtained from (5) and (2) were checked for consistency with the requirements of detailed balance by applying the operator to an equilibrium distribution function and verifying that the result was zero.

In the present calculations the deformation potential interaction was included for scattering with acoustic phonons and the Fröhlich interaction was included for scattering with longitudinal optical (LO) phonons (Conwell, 1967). The effects of these phonon scattering mechanisms on the $I(V)$ curve of the RTD are shown in Fig. 2. Acoustic phonon scattering has a nearly negligible effect on the $I(V)$ curve. The effect of LO phonon scattering is rather more pronounced, primarily in the reduction of the peak current. When both acoustic and LO phonons are included in the calculation, the resulting $I(V)$ curve is indistinguishable from that obtained with LO scattering only.

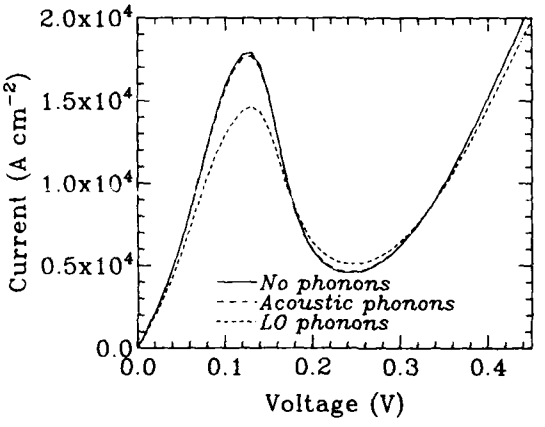


Fig. 2. Effect of semi-classical phonon-scattering operators on the $I(V)$ characteristic of a resonant-tunneling diode.

The effect of phonon scattering on the dynamic behavior of the resonant-tunneling diode was investigated by performing transient-response calculations (Frensley, 1986, 1987a) both with and without the collision operator. The particular transient event that was simulated was an instantaneous switching of the applied voltage from the peak of the $I(V)$ curve to the valley. The current through the RTD (averaged with respect to position within the structure) is illustrated in Fig. 3. Both curves show an initial peak and small oscillations around a generally exponential decay to the new steady state. Somewhat surprsingly, these oscillations are larger for the calculation including phonon scattering.

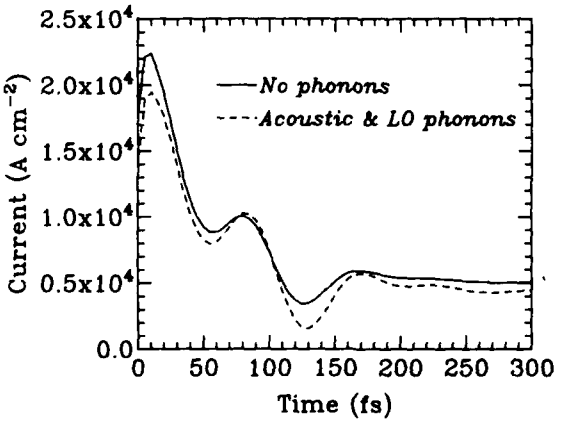


Fig. 3. Transient response of a resonant-tunneling diode with and without phonon scattering.

CONCLUSIONS

The present work represents an initial effort to include phonon scattering effects in a transport theory of tunneling devices. The diversity of results obtained from the simpler approximations to the collision operator clearly indicates that meaningful results (and reliable insights into the effects of stochastic processes on nanoelectronic devices) will only be obtained from realistic models of these phenomena.

ACKNOWLEDGEMENT

This work was supported in part by the Office of Naval Research and the U.S. Army Research Office.

REFERENCES

Caldeira, A.O. and A.J. Leggett (1983). Path integral approach to quantum Brownian motion. *Physica*, 121A, 587-616.

Chang, L.L., L. Esaki, and R. Tsu (1974). Resonant tunneling in semiconductor double barriers. *Appl Phys Lett.*, 24, 593-5.

Conwell, E.M. (1967). *High Field Transport in Semiconductors*. Academic Press, New York, ch. 5.

Frensley, W.R. (1986). Transient response of a tunneling device obtained from the Wigner function. *Phys Rev Lett.*, 57, 2853-6.

Frensley, W.R. (1987a). Wigner function model of a resonant-tunneling semiconductor device. *Phys Rev B*. Accepted for publication.

Frensley, W.R. (1987b). Quantum-transport calculation of the small-signal response of a resonant-tunneling diode. *Appl. Phys Lett*. Accepted for publication.

Kubo, R., M. Toda, and N. Hashitsume (1985). *Statistical Physics II. Nonequilibrium Statistical Mechanics* Springer-Verlag, Berlin, p. 55.

Levinson, I.B. (1970). Translational invariance in uniform fields and the equation of motion for the density matrix in the Wigner representation. *Soviet Physics JETP*, 30, 362-7.

Lin, J. and L.C. Chiu (1985). Quantum theory of electron transport in the Wigner formalism. *J. Appl Phys.*, 57, 1373-6.

Sollner, T.C.L.G., W.D. Goodhue, P.E. Tannenwald, C.D. Parker, and D.D. Peck (1983). Resonant tunneling through quantum wells at frequencies up to 2.5 THz. *Appl. Phys. Lett.*, 43, 588-90.

Wigner, E. (1932). On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, 40, 749-59.

Appendix D

"Boundary Conditions for Open Quantum Systems
Driven Far from Equilibrium"

[Please note that this is an incomplete draft manuscript. Before it is submitted for publication, several more issues will be developed, and related work by other researchers will be discussed.]

Boundary Conditions for Open Quantum Systems Driven Far from Equilibrium

William R. Frensley
Central Research Laboratories
Texas Instruments Incorporated
Dallas, Texas 75265

Abstract

Boundary conditions are formulated for kinetic models for open systems, in the sense of systems which can exchange conserved particles with their environment. Opening a system to particle flow violates the Hermiticity of the Liouville operator. If the open-system boundary conditions are time-reversible, exponentially growing (unphysical) solutions are introduced into the time-dependence of the density matrix. This problem is avoided by applying time-irreversible boundary conditions to the Wigner distribution function. These boundary conditions model the external environment as ideal particle reservoirs with properties analogous to those of a blackbody.

1 Introduction

The more active, and thus the more interesting, products of technology are systems which operate far from thermal equilibrium. An examination of a few examples of such systems shows that they are generally open, in the sense that they exchange matter with their environment. The present work examines some schemes by which open quantum systems (which are beginning to become technologically important in the context of microelectronics) may be effectively described at a kinetic level.

In the context of the present work, an "open system" is one which can exchange locally conserved particles with its environment. To define such a system we must regard it as occupying a finite region of space, and thus the exchange of particles must consist of a

current flowing through that surface which is taken to be the boundary of the system. It does not appear that the statistical physics of such a situation has been the subject of a close examination, apart from the traditional use of the grand canonical ensemble to define the equilibrium state [1]. There is, of course, a large body of work on quantum systems which are coupled to a reservoir so as to permit an exchange of energy [2,3,4,5,6]. Such analyses are more directed to the problem of damping (as seen in ohmic conduction) than to openness in the present sense. Much of the work in this area has been motivated by the development of optical technology [5,6], in which the distinction between openness and damping is unnecessary because the particles of interest are massless bosons.

To document the importance of open systems, let us consider some examples of active systems. Most practical engines (in the sense of machines which convert some form of energy into mechanical work) exchange matter with two or more reservoirs. To cite examples from an earlier technology we might consider the overshot water wheel [7], which operates between reservoirs of water at different gravitational potential, or the high pressure steam engine [8], which operates between its boiler and the atmosphere, reservoirs which differ greatly in their pressure and temperature. Conspicuously absent from a list of economically significant engines are systems which operate upon the Carnot model of a closed system in purely thermal contact with its reservoirs.

A technology of more current interest is electronics, whose systems are usually arranged such that a "power supply" maintains constant voltages (*i.e.*, chemical potentials for electrons) on two or more "buses" [9]. The "circuits" (such as logical gates or analog amplifiers) which perform the intended functions of the system are connected to, and conduct current between, the buses. Each bus is an electron reservoir, and the performance of the system's power supply is judged by how nearly these reservoirs approach the ideal behavior of no change in chemical potential (voltage) as particles are exchanged (current is drawn).

The example of electronics points out that the distinction between a closed and an open system depends upon how one chooses to partition the universe into the system of

interest and "everything else." (Such a partitioning is implicit in the analysis of every physical problem.) To demonstrate this point, let us examine the etymology of the term *circuit.* As used in the preceding paragraph, *circuit* means "an assemblage of electronic elements [10]," which is most often open with respect to electron flow. This usage of the term is now much more common among electrical engineers than the original meaning, "the complete path of an electric current including usually the source of electric energy [10]," which implies a closed system with respect to electron flow. It is no accident that the usage of the word *circuit* has evolved in this manner. Early in the development of electrical technology, a useful system (such as the electromagnetic telegraph [11]) was composed of at most a few topologically closed "circuits," and the closure of the current path was a central concern. As the complexity of electrical systems increased, the power supply and bus structure was developed. This provided a common segment for all the current paths, and the attention of the engineer focused on the remaining, "interesting," segment, that which contained the active devices (and the term *circuit* came to be applied to such a segment). However, by focusing on only a segment of the current path, one had to deal with an open system, rather than a closed one.

The physics of closed systems is certainly simpler than that of open systems, because closed systems obey global conservation laws, which open systems, in general, do not. In the well-established techniques of physical theory one often encounters artifices, usually in the form of periodic boundary conditions, which assure the closure of the theoretical model, if not of the system itself. The point of the present discussion is that it is frequently necessary to partition a complex system (which might reasonably be regarded as closed) into smaller components which, viewed individually, must be regarded as open. Thus, the more applied disciplines of the physical sciences must often deal at some level with the concept of an open system.

There are many established techniques for dealing with open systems in fields such as fluid dynamics, neutron transport, and electronics. All these fields are concerned with the transport of (usually) conserved particles. The transport phenomena are described by transport equations at a kinetic or hydrodynamic level which are either differential

or integro-differential equations. Such equations require boundary conditions and it is in these boundary conditions that the openness of a system is described. In the computation of the flow around an airfoil, one must supply "upstream" and "downstream" boundary conditions [12], or, more generally, sources and sinks. In electronics the connection to the external circuit is accomplished by some sort of contact. In solid-state electronics the most frequently used type of contact is the ohmic contact, an interface between a metallic conductor and (usually) a semiconductor which permits electrons to pass freely. Because the ohmic contact is a critical component of solid-state technology, most work on such interfaces has been directed toward their fabrication and characterization[13]. The representation of such contacts by boundary conditions has been a part of the analysis of semiconductor device problems since the begining of semiconductor technology [14,15]. The current practice in this field is discussed in detail by Selberherr [16].

2 Approaches to Open Quantum Systems

Tunneling devices form the most obvious class of open systems in which quantum effects are significant. The most common approach to describing the behavior of such systems is to invoke the formal theory of scattering and assume that one is dealing with wavefunctions which asymptotically approach traveling waves[17]. This is quite adequate to describe the steady states of an open system which is not subject to any other dissipative interactions. If there is dissipation, in the form of ohmic resistance for example, one should describe the system at a level which admits mixed states. However, even if the motion of the particles within the system is purely ballistic (no dissipation) problems are encountered when one attempts to apply the scattering approach to the evaluation of transient phenomena. The reason for this is that the boundary conditions (at any fixed position) for the Schroedinger equation which permit outgoing waves to pass through without reflection are nonlocal in time (or non-Markovain). To examine this, let us compare the Schroedinger equation to the simple wave equation, for which one *can* formulate nonreflecting boundary conditions.

For the simple wave equation,

$$\frac{\partial^2\phi}{\partial t^2} - c^2\frac{\partial^2\phi}{\partial x^2} = 0,$$

one must specify the value of $\partial\phi/\partial x$ at each end of the domain (Neumann boundary conditions). Fourier transforming with respect to both position and time, and solving for the Fourier transform of the gradient, we get the simple expression

$$ik = \pm i\omega/c.$$

The choice of sign determines which direction of propagation is permitted to pass without reflection. Inverting the Fourier transform, we obtain the boundary conditions (for the homogeneous case)

$$\frac{\partial\phi}{\partial x} = \pm\frac{1}{c}\frac{\partial\phi}{\partial t},$$

and these are easily implemented in practical computations because the time derivative of the wavefunction is readily available.

Let us consider applying this procedure to the Schroedinger equation (assuming for the moment that the potential at any boundary is zero). We find

$$ik = i\,(2im\omega/\hbar)^{\frac{1}{2}}\,.$$

If we attempt to invert the Fourier transform, the $\omega^{1/2}$ factor does not lead to a finite series of time derivatives; instead, it leads to an integral expression which depends upon the entire history of the system (that is, a memory term). If one wishes to evaluate the behavior of a model system over a finite domain in position and time, such a memory term is a great nuisance, although it appears that it may be adequately approximated over a limited range of energies [18]. It is not at all obvious that such a memory term is really necessary to describe the behavior of something like a tunneling device. Thus, it is desirable to possess an alternative model for the behavior of an open system which does not contain a memory term. Such a model will be Markovian, in the sense that the equation of motion will be a first-order differential equation with respect to time, and the present work is restricted to an examination of such models. Markovian open-system models may be formulated at a statistical (kinetic) level, and are not an approximation to

scattering theory, but make use of a different set of assumptions. These assumptions can also be viewed as the approximations by which a many-body problem is reduced to a more tractable form.

3 Quantum Kinetic Theory

A generally accepted approach to the problems of statistical physics is to begin with the general theory of many-body dynamics and to proceed by deductive reasoning to a formulation which provides an answer for the problem of interest [19]. The steps in this deductive chain necessarily involve the introduction of extra assumptions in the form of suitable approximations. One may loosely categorize the levels of approximation in terms of the independent variables which are required to specify the state of a system. The most detailed level is the fundamental many-body theory, which in principle requires a complete set of dynamical variables for each particle. This can be reduced to the kinetic level by restricting one's attention to one- or two-body properties (by truncating the BBGKY hierarchy of equations, for example). The removal from explicit consideration of other dynamical variables of the complete system, such as photon or phonon coordinates, may also be required. The kinetic theory is expressed in terms of distribution functions defined on a single-particle phase space, requiring one position and one momentum variable for each spatial dimension. (In the quantum case, this goes over to two arguments of the density operator.) The hydrodynamic level of approximation is obtained by making some assumption about the form of the distribution function with respect to momentum, and integrating over all of the momenta. Thus, the hydrodynamic theory is expressed in terms of densities which are functions of position only.

The approach taken in the present work is quite different from the conventional deductive approach. The objective is to identify the mathematical properties which are required of simple kinetic models of open systems. The procedure will be to construct small, spatially discretized models and to numerically explore their properties. The significance of the results must then be argued inductively.

In the kinetic level of description of a complex system, the effects of those degrees of freedom which are of less interest in a given problem are included implicitly in objects such as collision operators or effective interaction potentials. In the example of electronic devices such degrees of freedom should include electron coordinates outside the device but within the external circuit. They also include all excitations of the device material apart from the single-electron states (*e.g.*, the phonons). Thus, at this level the state of the system is described by a one-body density operator or distribution function. In general, this can be written as

$$\rho(x,x') = \sum_i w_i \langle x \mid i \rangle \langle i \mid x' \rangle, \tag{1}$$

where i labels a complete set of states and the w_i are real-valued probabilities for the system to be in state $\mid i\rangle$. Because we will be considering open systems in which the number of particles is not fixed, the usual convention for the normalization of ρ ($\langle i \mid i \rangle = 1$ and $\mathrm{Tr}\,\rho = 1$) is not useful. Instead, we will adopt a normalization convention such that $\rho(x,x)$ gives the actual particle density (in units of particles per cm^3, for example). Thus, the definition of the density operator is subtly changed from the probability distribution for an ensemble of identical single-particle systems to the occupation factor for a system composed of many noninteracting particles. One of the consequences of this is that Maxwell-Boltzmann statistics naturally follow from this picture. Therefore, the models to be developed are appropriate for dilute systems, and such effects as Fermi degeneracy or correlation effects must be incorporated in the form of correction terms.

For a system described by a simple single-particle Hamiltonian,

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + v(x), \tag{2}$$

the time evolution of the density matrix is given by the Liouville-von Neumann equation:

$$\begin{aligned} i\hbar\frac{\partial \rho}{\partial t} &= [H,\rho] \triangleq \mathcal{L}\rho \\ &= -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial x'^2}\right)\rho + [v(x) - v(x')]\,\rho, \end{aligned} \tag{3}$$

where H is the Hamiltonian and $\mathcal{L}$ is the Liouville superoperator. The simplest approach to modeling the behavior of open systems is to apply the Liouville equation to a finite

spatial domain representing the system of interest and to apply boundary conditions which model the openness of the system. The difficulties and ultimate success of this approach involve the effect that such boundary conditions have upon the properties (particularly the eigenvalue spectrum) of the Liouville superoperator. Zwanzig [20] has presented an excellent discussion of the properties of superoperators (or tetradics). However, the present analysis requires a somewhat different group of expressions, so the subject will be developed here.

The density operators which represent the state of a statistically mixed system themselves form a linear vector space analogous to the space of pure quantum states represented by wavefunctions. A linear combination of density operators might be used to describe the results of superposing two partially polarized beams of particles, for example. Anything which generates linear transformations on a density operator [such as the right-hand side of the Liouville equation (3)] is a superoperator. In a finite, discrete system with N states, a wavefunction will be a vector (a singly-indexed object) with N elements, a density operator will be a matrix (a doubly-indexed object) with N^2 elements, and a superoperator will be a tetradic (a quadruply-indexed object) with N^4 elements. Superoperators are isomorphic to ordinary operators, but to define concepts such as Hermiticity or unitarity of superoperators, we must have a definition for the inner product of two ordinary operators. The simplest definition is

$$\langle A \| B \rangle = \mathrm{Tr}(A^\dagger B), \tag{4}$$

where A and B are operators and the notation $\langle \| \rangle$ is introduced to indicate expressions in the linear space of operators. It is easily shown that this satisfies the axioms [21] defining an inner product on a complex vector space. Then a Hermitian superoperator $\mathcal{H}$ satisfies

$$\langle A \| \mathcal{H} B \rangle = \langle \mathcal{H} A \| B \rangle \tag{5}$$

and a unitary superoperator $\mathcal{U}$ satisfies

$$\langle \mathcal{U} A \| \mathcal{U} B \rangle = \langle A \| B \rangle. \tag{6}$$

Superoperators are usually derived from ordinary quantum observable operators by forming the commutator or anti-commutator with the operator being acted upon. For an

operator C let us denote these superoperators

$$\mathcal{C}_{(-)}A = CA - AC, \tag{7}$$

$$\mathcal{C}_{(+)}A = \tfrac{1}{2}\left[CA + AC\right]. \tag{8}$$

If C is Hermitian ($C^\dagger = C$) the Hermiticity of $\mathcal{C}_{(-)}$ and $\mathcal{C}_{(+)}$ follow immediately:

$$\begin{aligned}\langle \mathcal{C}_{(-)}A \| B\rangle &= \mathrm{Tr}\left[(CA - AC)^\dagger B\right] = \mathrm{Tr}\left(A^\dagger C^\dagger B - C^\dagger A^\dagger B\right) \\ &= \mathrm{Tr}\left(A^\dagger CB - CA^\dagger B\right) = \mathrm{Tr}\left(A^\dagger CB - A^\dagger BC\right) \\ &= \mathrm{Tr}\left[A^\dagger (CB - BC)\right] = \langle A \| \mathcal{C}_{(-)}B\rangle,\end{aligned} \tag{9}$$

and similarly for $\mathcal{C}_{(+)}$. The Hermiticity (or lack thereof) of the Liouville superoperator is the critical issue in formulating a kinetic model of open systems.

4 Time-Reversible Open System Model

To describe the behavior of an open system, we will consider an approach in which the spatial domain is considered to be finite, corresponding to the extent of the system, and boundary conditions are applied which permit particles to pass into and out of the system. The first model we will consider employs time-reversible boundary conditions which are plausible, but which we will ultimately see to be unphysical [22]. The reason for examining this model is that it helps to define the conditions that a physically reasonable open-system model must display.

To provide the motivation for the first model, let us consider a spatially uniform particle gas of infinite extent, $-\infty < x < \infty$, and take the open system to be the finite region $0 \leq x \leq l$. The thermal equilibrium density matrix for a uniform gas may be obtained by integrating the Bloch equation [23]

$$\partial \rho_{\mathrm{eq}}/\partial \beta = -H\rho_{\mathrm{eq}}. \tag{10}$$

The solution $\rho_{f\mathrm{eq}}$ (for free particles in equilibrium)

$$\rho_{f\mathrm{eq}}(x, x') = \sqrt{\frac{m}{2\pi\hbar^2\beta}} \exp\left[-\left(\frac{m}{2\hbar^2\beta}\right)(x - x')^2 + \beta\mu\right], \tag{11}$$

where the normalization is such that $\rho_{\mathrm{feq}}(x,x)$ gives the number of particles per unit length and μ is the chemical potential. Now if we arbitrarily impose boundaries along the lines $x = 0$, $x = l$, $x' = 0$, and $x' = l$, what boundary conditions would ρ_{feq} satisfy? Note that the dependence is only upon $(x - x')$, so that $\partial\rho/\partial x = -\partial\rho/\partial x'$. Thus, in this particular case ρ obeys the homogeneous boundary condition

$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial x'}\right)\rho\Big|_{\mathrm{boundary}} = 0. \tag{12}$$

Is (12) the appropriate boundary condition for a general open system? Let us explore some of its consequences. Suppose at time $t = 0$ we apply a uniform force field F to the particle gas. The solution to the Liouville equation (3) over the infinite domain and with initial condition (11) describes an accelerating gas and is given by

$$\rho_{\mathrm{acc}}(x,x';t) = \rho_{\mathrm{feq}}(x,x')\exp\left[\left(\frac{iFt}{\hbar}\right)(x - x')\right]. \tag{13}$$

Now ρ_{acc} also obeys (12), so it is also the solution to (3) over the finite domain subject to boundary condition (12).

A more general consequence of boundary condition (12) is that the particle densities at the boundaries, $\rho(0,0)$ and $\rho(l,l)$, remain constant as the density matrix evolves with time. To demonstrate this, note that we can factor the hyperbolic operator in the Liouville equation (3) derived from the kinetic energy terms as

$$\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial x'^2} = \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial x'}\right). \tag{14}$$

The boundary condition assures that the second factor in (14) is zero along the boundaries, and along the diagonal the potential term is zero. Thus, $\partial\rho(0,0)/\partial t = 0$ and $\partial\rho(l,l)/\partial t = 0$, the boundary densities do not change. This might be interpreted as the behavior of a large reservoir with a fixed particle density (or fixed pressure if the temperature is also fixed). Thus, the boundary condition (12) provides a plausible model for an open system.

In fact, the Liouville equation (3) subject to the boundary condition (12) generates an unphysical solution in the form of exponentially growing particle densities when it is applied to more general potentials which do not have the symmetry of the uniform field

[22]. The nature of the time-dependent solutions (whether they are growing, decaying, or oscillatory) depends upon the eigenvalue spectrum of the Liouville superoperator (the definition of which requires both the differential operator and the boundary conditions). The problem with the growing densities (and ultimately the identification of the correct model) is a consequence of opening the system, which violates the Hermiticity of the Hamiltonian operator and of the Liouville superoperator. Recall the proof [24] of the Hermiticity of the Hamiltonian (2). It proceeds by invoking Green's identity to transpose the Laplace operator, which leaves a surface term. The precise expression is

$$\int_V (H - H^\dagger)\, d^3\mathbf{x} = \frac{\hbar}{i} \int_S \mathbf{j} \cdot d^2\mathbf{s}, \tag{15}$$

where V refers to the volume of the domain, S is its surface, and $\mathbf{j}$ is the current-density operator. One maintains the Hermiticity of the Hamiltonian by choosing basis functions for which the the surface integral is identically zero: states well localized within the domain, and stationary scattering states (or periodic boundary conditions) for which the incoming and outgoing currents cancel. Because the total number of particles in an open system can change in response to externally imposed conditions, such a basis set is too restrictive.

The violation of the Hermiticity of the Liouville superoperator follows directly from that of the Hamiltonian. This leads to eigenvalues of the Liouville superoperator which have nonzero imaginary parts, leading to real exponential behavior in the time dependence of ρ. There is a different, but related, situation in which such behavior is desired (and in fact necessary): the case in which motion is damped by dissipative interactions. A specific example is the ohmic dissipation associated with normal electrical conduction in solids. Dissipation has been a very widely studied phenomenon, at all levels from fundamental statistical physics [2] to the engineering properties of specific materials [25]. In the spirit of the present treatment, however, let us take an extremely simple model of dissipation, which we will use to study the relationship between openness and damping. This model is simple Brownian motion as described by the Fokker-Planck or Kramers equation [26,27]. It is classically valid in the limit that the particles of interest are weakly coupled to an ideal reservoir. Caldeira and Leggett [27] have studied the quantum-mechanical derivation

of this equation and have shown it to be valid at higher temperatures. In terms of ρ the Fokker-Planck equation may be written

$$\frac{\partial \rho}{\partial t} = \frac{1}{i\hbar}\mathcal{L}\rho - \mathcal{D}\rho, \tag{16}$$

where $\mathcal{D}$ is a damping superoperator. The Fokker-Planck expression for $\mathcal{D}$ is

$$\mathcal{D}\rho = \gamma\left[\frac{(x-x')}{2}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)\rho + \frac{m}{\hbar^2\beta}(x-x')^2\rho\right], \tag{17}$$

where γ is the damping rate. The first term in (17) describes dissipation and corresponds to a frictional force equal to γp, where p is the linear momentum. The second term describes the thermal fluctuations. An important property of $\mathcal{D}$ is that $(\mathcal{D}\rho)(x,x) = 0$, which is required for consistency with the continuity equation. $\mathcal{D}$ will be used below to add dissipative interactions to our open-system models.

To explore the eigenvalue spectrum of the present open-system model and those which will be investigated later, let us consider a finite-difference approximation to the Liouville equation (3) which reduces $\mathcal{L}$ to a finite matrix whose eigenvalues may be readily computed. The position coordinates x will be taken to be elements of a uniformly spaced mesh: $\{\, x_j \mid x_j = j\Delta \text{ for } j = 1, 2, \ldots, N \,\}$. The dependent quantities such as the wavefunction and density matrix then take on discrete values also, which will be denoted by $\psi_j = \psi(x_j)$, and $\rho_{ij} = \rho(x_i, x_j)$. Using the simple finite-difference approximation $(\partial^2\psi/\partial x^2)_i = (\psi_{i-1} - 2\psi_i + \psi_{i+1})/\Delta^2$, the Hamiltonian (2) becomes

$$H_{ij} = \frac{\hbar^2}{2m\Delta^2}(2\delta_{ij} - \delta_{i-1,j} - \delta_{i+1,j}) + v_i\delta_{ij}, \tag{18}$$

for i, j not on one of the boundaries. To incorporate the boundary conditions, it is best to think of adding an additional mesh point at each end of the domain (points x_0 and x_{N+1}), and specify the value of the wavefunction on those points. For example, to apply the homogeneous Dirichlet conditions for a particle in a box, we would set $\psi_0 = 0$ and $\psi_{N+1} = 0$. Inserting these conditions into (18) completely defines the matrix H_{ij} for $1 \le i, j \le N$. Similarly, if we wanted to apply Neumann conditions $\partial\psi/\partial x = 0$, we would set $\psi_0 = \psi_1$.

Writing the Liouville equation (3) on the finite-difference basis gives

$$i\hbar\,(\partial\rho/\partial t)_{ij} = \mathcal{L}_{ij;kl}\rho_{kl}, \tag{19}$$

where the tetradic nature of $\mathcal{L}$ is made explicit. This discrete representation of $\mathcal{L}$ may be derived from (18) and is

$$\mathcal{L}_{ij;kl} = \frac{\hbar^2}{2m\Delta^2}\left(-\delta_{i-1,k}\delta_{jl} - \delta_{i+1,k}\delta_{jl} + \delta_{ik}\delta_{j-1,l} + \delta_{ik}\delta_{j+1,l}\right) + (v_i - v_j)\,\delta_{ik}\delta_{jl}. \tag{20}$$

Again, the elements adjacent to a boundary require special attention.

To evaluate the eigenvalues of $\mathcal{L}$ and other superoperators, we must map the tetradic onto an ordinary matrix, so that conventional eigenvalue algorithms may be applied. To do so for the finite, discrete case, we may map the density matrix ρ onto a singly subscripted vector of dimension N^2 by $\rho_{ij} \rightarrow \rho_m$ with $m = (i-1)N + j$. (This kind of expression is well known to computer programmers as the means by which matrices are mapped into the linear address space of a computer.) Note that with this mapping the inner product between two operators (4) becomes the ordinary inner product between two vectors. The mapping of the tetradic $\mathcal{L}$ onto an $N^2 \times N^2$ matrix follows immediately. The matrix representing $\mathcal{L}$ was actually constructed for $N = 8$ (resulting in a 64×64 matrix for $\mathcal{L}$) using the potential illustrated in Fig. 1. The first case considered was a closed system with no damping. This model was obtained by simply applying the particle-in-a-box (homogeneous Dirichlet) boundary conditions to the Liouville operator (20). The eigenvalue spectrum which results is shown in Fig. 2(a). All of the eigenvalues are purely real, as expected from a Hermitian superoperator.

In the second case the model system is taken to be closed, but damped. The Fokker-Planck damping operator (17) may be written in discretized form as

$$\mathcal{D}_{ij;kl} = \frac{\gamma m\Delta^2}{\hbar^2\beta}(i-j)^2\delta_{ik}\delta_{jl} + \frac{\gamma\Delta}{2} \times \begin{cases} (i-j)[2\delta_{ik}\delta_{jl} - \delta_{i-1,k}\delta_{jk} - \delta_{ik}\delta_{j+1,l} & \text{for } i > j \\ (j-i)[2\delta_{ik}\delta_{jl} - \delta_{i+1,k}\delta_{jk} - \delta_{ik}\delta_{j-1,l} & \text{for } i < j \end{cases}. \tag{21}$$

This form preserves the important properties of $\mathcal{D}$. To illustrate the effect of dissipation on the spectrum of $(\mathcal{L} - i\hbar\mathcal{D})$, the zero temperature limit ($\beta \rightarrow \infty$) was taken and the

damping constant $\gamma = 0.01 \times (\hbar/2m\Delta^2)$ was used. The resulting eigenvalue spectrum is shown in Fig. 2(b). Negative imaginary parts have been introduced into all the eigenvalues (except possibly one eigenvalue which is equal to zero within the numerical roundoff error). These negative imaginary parts lead to damped motion, as expected.

Now with this background we can consider the case of the open-system boundary conditions (12). The simplest finite-difference approximation for the condition (12) is

$$\left(\frac{\partial\rho}{\partial x}+\frac{\partial\rho}{\partial x'}\right)_{ij}=\left[\frac{1}{\Delta}\left(\rho_{i+1,j}-\rho_{ij}\right)+\frac{1}{\Delta}\left(\rho_{ij}-\rho_{i,j-1}\right)\right]=\frac{1}{\Delta}\left(\rho_{i+1,j}-\rho_{i,j-1}\right)=0, \qquad (22)$$

for i or j equal to 1 or N. Thus, the open-system Liouville superoperator $\mathcal{L}^{(\mathrm{or})}$ (for open system, reversible) is obtained by inserting boundary values $\rho_{i0} = \rho_{i+1,1}$ and $\rho_{N+1,j} = \rho_{N,j-1}$ (and the expressions obtained by transposing the indices) into (20). For the sake of completeness, let us write down the elements of $\mathcal{L}^{(\mathrm{or})}$ which are affected by the boundary conditions:

$$\begin{aligned}
\mathcal{L}^{(\mathrm{or})}_{i1;kl} &= \frac{\hbar^2}{2m\Delta^2}\left(-\delta_{i-1,k}\delta_{1,l}+\delta_{ik}\delta_{2,l}\right)+\left(v_i-v_1\right)\delta_{ik}\delta_{1,l}.\\
\mathcal{L}^{(\mathrm{or})}_{1j;kl} &= \frac{\hbar^2}{2m\Delta^2}\left(-\delta_{2,k}\delta_{jl}+\delta_{1,k}\delta_{j-1,l}\right)+\left(v_1-v_j\right)\delta_{1,k}\delta_{jl}.\\
\mathcal{L}^{(\mathrm{or})}_{iN;kl} &= \frac{\hbar^2}{2m\Delta^2}\left(-\delta_{i+1,k}\delta_{Nl}+\delta_{ik}\delta_{N-1,k}\right)+\left(v_i-v_N\right)\delta_{ik}\delta_{Nl}.\\
\mathcal{L}^{(\mathrm{or})}_{Nj;kl} &= \frac{\hbar^2}{2m\Delta^2}\left(-\delta_{N-1,k}\delta_{jl}+\delta_{Nk}\delta_{j+1,l}\right)+\left(v_N-v_j\right)\delta_{Nk}\delta_{jl}.
\end{aligned}$$

The non-Hermiticity of $\mathcal{L}^{(\mathrm{or})}$ follows from these expressions. For example, $\mathcal{L}^{(\mathrm{or})}_{i,1;i-1,1} = -\hbar^2/(2m\Delta^2)$, but $\mathcal{L}^{(\mathrm{or})}_{i-1,1;i,1} = 0$. What has happened is that boundary conditions caused elements of $\mathcal{L}$ to be canceled in a way which breaks the Hermitian symmetry. The resulting eigenvalue spectrum is plotted in Figure 3(a). The non-Hermiticity of $\mathcal{L}^{(\mathrm{or})}$ leads to some eigenvalues with nonzero imaginary parts. It is apparent that these eigenvalues occur in complex conjugate pairs, with both positive and negative imaginary parts present. This is a consequence of the time-reversal symmetry of both the Liouville equation and the open-system boundary conditions (12). The eigenvalues with positive imaginary parts produce growing exponential solutions to the Liouville equation, which would prevent any approach to steady state. This open-system model is thus physically unacceptable.

One might speculate that the problem of growing solutions could be due to the absence of damping in the model. To test this, let us add in the Fokker-Planck damping term (21) as we did for the closed-system model. With the same damping constant [$\gamma = 0.01 \times (\hbar/2m\Delta^2)$] as before, the resulting eigenvalue spectrum for $(\mathcal{L}^{(\mathrm{or})} - i\hbar\mathcal{D})$ is that shown in Figure 3(b). The addition of damping clearly does not solve the stability problem because it does not remove the positive imaginary parts. In fact, a larger damping constant does lead to a stable model, as shown in Figure 3(c), where $\gamma = 0.03 \times (\hbar/2m\Delta^2)$ was used. All the eigenvalues now have negative imaginary parts, except for a doubly degenerate eigenvalue at zero (which must be present because of the invariance of ρ_{11} and ρ_{NN}).

Thus modeling an open system by applying the boundary conditions (12) will work only if the rate of damping within the system is sufficiently large (or, for the case of electron transport, if the mobility is sufficiently low). The minimum acceptable damping rate depends upon the magnitude of the imaginary parts of the eigenvalues of $\mathcal{L}^{(\mathrm{or})}$ for the undamped system, which in turn depends upon the form of the potential. In fact, the potential of Figure 1 was chosen because it produces larger imaginary parts than potentials with greater symmetry. All this adds up to a very unsatisfactory formulation for an open-system model. The problems may be traced to the time-reversal symmetry of the boundary conditions. To obtain a proper formulation, this symmetry must be broken.

5 Irreversible Open-System Model

To provide a physical motivation for the idea that openness necessarily involves time-irreversibility, let us consider another example system drawn from electronic technology, the vacuum thermionic device ("vacuum tube" or "valve") [28]. These devices were made by introducing two or more metallic electrodes into a vacuum through which electrons could be transported without dissipation. When a voltage was applied between anode and cathode (and the cathode heated to thermally excite electrons into the vacuum), a nonequilibrium steady state would be established with a nonzero current flowing. Such a nonequilibrium steady state cannot be established in a reversible (or Hamiltonian) system.

Consider what would happen if a population of electrons were introduced into some sort of trapping potential in ultrahigh vacuum. The system would effectively be closed, and the motion of the electrons would consist of periodic (thus, reversible) orbits. Of course what happened in the case of the thermionic vacuum tube is that electrons were accelerated by the electrostatic field until they impacted the anode, where they lost their kinetic energy to collisions with the electrons in the metal. Thus, the power was dissipated as heat. However, we can infer a much broader principle from this device: Making contact to a system in such a way as to permit particles to enter and leave (opening the system) in itself introduces irreversibility into the behavior of the system.

Now, if the openness of the system is to be modeled by boundary conditions applied to the system, these boundary conditions must themselves be time-irreversible. A physically reasonable way to achieve such irreversibility is to distinguish between particles moving into and out of the system. It is then reasonable to expect that the distribution of particles flowing into the system depends only upon the properties of the reservoirs to which the system is connected, and that the distribution of particles flowing out of the system depends only upon the state of the system. This picture leads to a fully acceptable model of an open system.

To implement boundary conditions which distinguish between patricles flowing into and out of a system, we must reexpress the Liouville equation (3) in terms of the classical phase space (q,p), where q in this case corresponds to the position x and p is the momentum. This is naturally done by the Wigner-Weyl transformation, which transforms the density operator $\rho(x,x')$ into the Wigner distribution function $f(q,p)$ [29,30]. For the present purposes, the Wigner-Weyl transformation consists of a change of independent coordinates to the diagonal and cross-diagonal coordinates:

$$\begin{aligned} q &= \tfrac{1}{2}(x+x'), \\ r &= x-x', \end{aligned} \tag{23}$$

followed by a Fourier transformation with respect to r. (The variables q and r are often referred to as "center of mass" and "relative" coordinates, respectively. I feel that this is a

misleading terminology, because it gives the incorrect impression that one is dealing with a two-body problem.) The variables x and x' may be expressed in terms of q and r by

$$\begin{aligned} x &= q + \tfrac{1}{2}r, \\ x' &= q - \tfrac{1}{2}r. \end{aligned} \tag{24}$$

Thus the Wigner distribution can be expressed as

$$f(q,p) = \int_{-\infty}^{\infty} dr\, \rho(q + \tfrac{1}{2}r, q - \tfrac{1}{2}r)\, e^{-ipr/\hbar}. \tag{25}$$

The Liouville equation becomes

$$\frac{\partial f}{\partial t} = -\frac{p}{m}\frac{\partial f}{\partial q} - \frac{1}{\hbar}\int_{-\infty}^{\infty} \frac{dp'}{2\pi\hbar} V(q, p - p') f(q, p'), \tag{26}$$

where the kernel of the potential operator is given by

$$V(q,p) = 2\int_0^{\infty} dr\, \sin(pr/\hbar)\left[v\left(q + \tfrac{1}{2}r\right) - v\left(q - \tfrac{1}{2}r\right)\right]. \tag{27}$$

This nonlocal potential is the means by which interference between alternative paths enters the Wigner-function formalism. The remaining part of the Liouville equation, the drift (or streaming, or advection) term, is exactly the same as the corresponding term of the classical Liouville equation with force F:

$$\frac{\partial f_{\rm cl}}{\partial t} = -\frac{p}{m}\frac{\partial f_{\rm cl}}{\partial q} - F\frac{\partial f_{\rm cl}}{\partial p}. \tag{28}$$

The correspondence between the classical and quantum drift terms will be exploited in defining the open-system boundary conditions.

To address the question of boundary conditions, first note that in the Wigner-Weyl representation, the Liouville equation (26) is of first order with respect to q and does not contain derivatives with respect to p. Thus, we must supply one boundary value (at $q = 0$ or $q = l$) for each possible value of p. There is no need to supply boundary conditions at any limiting values of p. The kinds of boundary conditions which are appropriate are illustrated in Figure 4. To implement the picture described above, that the particles entering the device depend only upon the state of the reserviors and that the particles

leaving the device depend only upon the state of the device, we should apply the boundary conditions illustrated in Figure 4(c). That is, we set

$$f(0,p)|_{p>0} = f^{(l)}_{\rm bdy}(p),$$
$$f(l,p)|_{p<0} = f^{(r)}_{\rm bdy}(p), \tag{29}$$

where $f^{(l)}_{\rm bdy}$ is the distribution function of the reservoir to the left of the system and $f^{(r)}_{\rm bdy}$ is the distribution function of the reservoir to the right. These boundary conditions are not invariant under time-reversal, because time-reversal would change the problem of Fig. 4(c) into that of Fig. 4(d).

To investigate the eigenvalue spectrum of the Liouville operator subject to the boundary conditions (29) (which we will refer to as $\mathcal{L}^{(oi)}$ for open-system, irreversible), we again construct a small discrete model. The position variable q will take the same set of discrete values that x did in the previous section: $\{\, q_j \mid q_j = j\Delta_q \,\text{for}\, j = 1, 2, \ldots, N_q \,\}$. The values of p are also restricted to a discrete set: $\{\, p_k \mid p_k = (\pi\hbar/\Delta_q)[(k - \frac{1}{2})/N_p - \frac{1}{2}] \,\text{for}\, k = 1, 2, \ldots, N_p \,\}$. The mesh spacing in the p direction is thus $\Delta_p = (\pi\hbar)/(N_p\Delta_q)$. The choice of the discrete values for p follows from a desire to avoid the point $p = 0$ and the need to satisfy a Fourier completeness relation, which will be discussed later. For the purposes of the present discussion, the Liouville superoperator (26) will be separated into two terms:

$$\mathcal{L}^{(oi)} = i\hbar\mathcal{T} + i\hbar\mathcal{V}, \tag{30}$$

where $\mathcal{T}$ is the superoperator derived from the kinetic energy term of the Hamiltonian:

$$\mathcal{T}f = -\frac{p}{m}\frac{\partial f}{\partial q}, \tag{31}$$

and $\mathcal{V}$ is the superoperator derived from the potential term:

$$(\mathcal{V}f)(q,p) = -\frac{1}{\hbar}\int \frac{dp'}{2\pi\hbar} V(q, p - p')f(q,p'). \tag{32}$$

The discrete version of the potential term is readily defined. First, we define a discrete potential kernel:

$$V_{jk} = \frac{2}{N_p}\sum_{j'=1}^{N_q/2} \sin\left(\frac{2k\Delta_p j'\Delta_q}{\hbar}\right)(v_{j+j'} - v_{j-j'}), \tag{33}$$

where j indexes position q, and k indexes momentum p. [Notice that (33) invokes values of v_j which are outside the domain $\{ q_j | j = 1 \ldots N_q \}$. This expresses the nonlocality of quantum phenomena and is one way in which the environment of an open system influences the system's behavior. The values which one assumes for v_j, where $j \leq 0$ or $j > N_q$, depend upon the nature of the environment. If ideal reservoirs are assumed, then setting these values equal to the potential at the appropriate boundary appears to be an adequate procedure.] The elements of $\mathcal{V}$ are then:

$$\mathcal{V}_{jk;j'k'} = -\delta_{jj'} V_{j,(k-k') \bmod N_p} / \hbar. \tag{34}$$

Note that the elements of $\mathcal{V}$ are real and that $\mathcal{V}_{jk;j'k'} = -\mathcal{V}_{j'k';jk}$ so $(i\hbar\mathcal{V})$ is an imaginary Hermitian superoperator.

The boundary conditions (29) affect the form of the drift term $\mathcal{T}$ because they determine the proper finite-difference form for the gradient. On a discrete mesh a first derivative $(\partial f/\partial q)(q_j)$ can be approximated by either a left-hand difference

$$\left(\frac{\partial f}{\partial q}\right)_{\text{left}} (q_j) = \frac{f(q_j) - f(q_{j-1})}{\Delta_q}, \tag{35}$$

or a right-hand difference

$$\left(\frac{\partial f}{\partial q}\right)_{\text{right}} (q_j) = \frac{f(q_{j+1}) - f(q_j)}{\Delta_q}. \tag{36}$$

(There is also a centered difference form, $[f(q_{j+1}) - f(q_{j-1})]/2\Delta_q$, which has poor stability properties when used to approximate a drift term.) The boundary conditions determine which of the above difference forms must be used simply because one or the other will not couple the boundary value into the domain. Again, let us imagine that the boundary conditions (29) are implemented by fixing the value of f on mesh points just outside the domain:

$$\begin{aligned} f_{0,k} &= f^{(\text{l})}_{\text{bdy}\,k} \quad \text{for } p_k > 0, \\ f_{N_q+1,k} &= f^{(\text{r})}_{\text{bdy}\,k} \quad \text{for } p_k < 0. \end{aligned} \tag{37}$$

This scheme is illustrated in Figure 5. Consider $p_k > 0$. The boundary conditions are specified for q_0, and if this value is to be coupled into the domain, we must use the left-hand difference formula (35) for the gradient at q_1. Consistency then requires that we

use the left-hand difference for all q_j (for $p_k > 0$). Similarly, we must use the right-hand difference (36) for $p_k < 0$. In the context of hydrodynamic calculations such a difference scheme is called an "upwind" or "upstream" difference and is known to enormously enhance the stability of a computation [31]. It has also been used in neutron transport calculations at the kinetic (phase space) level [32]. The elements of $\mathcal{T}$ are thus:

$$\mathcal{T}_{jk;j'k'} = -\frac{p_k}{m\Delta_q}\delta_{kk'} \times \begin{cases} \delta_{j+1,j'} - \delta_{j,j'} & \text{for } p_k < 0 \\ \delta_{j,j'} - \delta_{j-1,j'} & \text{for } p_k > 0 \end{cases} . \tag{38}$$

The terms $\mathcal{T}_{1,k;0,k}$ and $\mathcal{T}_{N_q,k;N_q+1,k}$ couple to the fixed boundary values of f and are thus the coefficients of inhomogeneous terms and are not strictly elements of $\mathcal{T}$. (In particular, these terms are not included in the eigenvalue calculation because eigenvalues are properties of homogeneous linear operators.)

The eigenvalue spectrum for $\mathcal{L}^{(\mathrm{oi})}$ constructed from (30), (34), and (38) is shown in Figure 6. The potential of Fig. 1 was used, with $N_q = 8$ and $N_p = 8$. All the eigenvalues of $\mathcal{L}^{(\mathrm{oi})}$ have negative imaginary parts. Thus, the time-dependence of f contains only decaying exponentials, so the model is stable. It has been successfully employed to model the behavior of resonant-tunneling semiconductor devices [33,34].

The stability of this model follows from the boundary conditions (29) and does not depend upon discretization. To demonstrate this, let us consider the expectation value of $(\mathcal{L}^{(\mathrm{oi})}/i\hbar)$ with respect to an arbitrary distribution f: $\langle f\|(\mathcal{L}^{(\mathrm{oi})}/i\hbar)\|f\rangle$. If we demonstrate that this is nonpositive for any f, we will have shown that no eigenvalue of $(\mathcal{L}^{(\mathrm{oi})}/i\hbar)$ has a positive real part, because the operator itself is purely real. In the Wigner-Weyl representation the operator inner product (4) becomes simply [35]

$$\langle f\|g\rangle = \frac{1}{2\pi\hbar}\int dq \int dp\, f(q,p)g(q,p). \tag{39}$$

The expectation value can be rewritten

$$\langle f\|(\mathcal{L}^{(\mathrm{oi})}/i\hbar)\|f\rangle = \langle f\|\mathcal{T}\|f\rangle + \langle f\|\mathcal{V}\|f\rangle = \langle f\|\mathcal{T}\|f\rangle, \tag{40}$$

because $\langle f\|\mathcal{V}\|f\rangle = 0$ from the antisymmetry of $\mathcal{V}$. For the homogeneous problem the boundary conditions are $f(0,p) = 0$ for $p > 0$ and $f(l,p) = 0$ for $p < 0$. With this we can

integrate the expectation value for $\mathcal{T}$ and simplify it to obtain

$$\begin{aligned} \langle f \| \mathcal{T} \| f \rangle &= \frac{1}{4\pi\hbar m} \left[\int_{-\infty}^{\infty} p f^2(0,p)\, dp - \int_{-\infty}^{\infty} p f^2(l,p)\, dp \right] \\ &= \frac{1}{4\pi\hbar m} \left[\int_{-\infty}^{0} p f^2(0,p)\, dp - \int_{0}^{\infty} p f^2(l,p)\, dp \right] \leq 0. \end{aligned} \tag{41}$$

Thus, the stability of the solutions to the Liouville equation using $\mathcal{L}^{(\mathrm{oi})}$ follow from the boundary conditions alone. The physical significance of this argument is that the particles in an open system will eventually escape and the density will approach zero if there is no inward current flow from the environment. There is, however, a possible exception to this statement. If the potential has a local minimum within the system deep enough to create one or more bound states, any particles in those states will not escape. Such states should lead to eigenvalues of $\mathcal{L}^{(\mathrm{oi})}$ which are equal to zero, although in an open system with finite extent and finite potential depth there is presumably a nonzero rate of escape from any such state.

Let us examine how this open system model can be used. Consider first the evaluation of a nonequilibrium steady state. Note that if there are no bound states, all the eigenvalues of $\mathcal{L}^{(\mathrm{oi})}$ are nonzero (see Fig. 6). Thus, $\mathcal{L}^{(\mathrm{oi})}$ is a nonsingular operator and its inverse exists. Therefore, the solution to the steady-state equation, $\mathcal{L}^{(\mathrm{oi})} f = 0$, exists and is unique. The boundary conditions are inhomogeneous, leading to nonzero terms on the right-hand side of this equation. The steady-state Wigner function $f^{(\mathrm{dc})}$ (for "direct current") can thus be expressed as

$$f^{(\mathrm{dc})} = \sum_{k'|p_{k'}<0} \left(\mathcal{L}^{(\mathrm{oi})}/i\hbar \right)^{-1}_{j,k;N_q,k'} \frac{p_{k'}}{m\Delta_q} f^{(\mathrm{r})}_{\mathrm{bdy}\,k'} - \sum_{k'|p_{k'}>0} \left(\mathcal{L}^{(\mathrm{oi})}/i\hbar \right)^{-1}_{j,k;1,k'} \frac{p_{k'}}{m\Delta_q} f^{(\mathrm{l})}_{\mathrm{bdy}\,k'}. \tag{42}$$

In practice one does not construct the complete inverse $\mathcal{L}^{(\mathrm{oi})^{-1}}$, but instead solves the discrete linear system with a specific set of boundary values. For convenience in the discussion of the mathematical properties of this model which follows, let us write down the complete irreversible open-system Liouville equation in the discrete approximation:

$$\frac{\partial f_j, k}{\partial t} = -\frac{p_k}{m\Delta_q} \times \left\{ \begin{array}{ll} f_{j+1,k} - f_{j,k} & \text{for } p_k < 0 \\ f_{j,k} - f_{j-1,k} & \text{for } p_k > 0 \end{array} \right\} - \frac{1}{\hbar} \sum_{k'} V_{j;k,k'} f_{j,k'}, \tag{43}$$

where the notation $V_{j;k,k'} = V_{j,(k-k')\mathrm{mod}N_p}$ is introduced to shorten the expressions to be derived from the discrete Liouville equation. Solutions of this system of equations may be used to evaluate the steady-state current versus applied force (the current-voltage curve in the case of electron devices), or they can provide the initial state for a computation of the transient response of the open system [33]. The response to a small periodic perturbation (the small-signal ac response) may also be evaluated by expanding the Liouville superoperator in a perturbation series, and applying the perturbation to the steady-state solution [34].

Compare the present approach to the most commonly studied problem in transport theory, transport in a spatially homogeneous system with a uniform driving field (as is done to evaluate transport coefficients such as mobilities) [36,25]. This generates a mathematically homogeneous problem, and the solution corresponds to the null space of that superoperator which appears in the transport equation [37]. Thus, the superoperator must be singular and, if the transport equation is linear, the solution is not unique (the total density is not determined). What the present model demonstrates is that a proper formulation of transport through a spatially inhomogeneous system leads to a mathematically inhomogeneous problem, which is in many respects a good deal simpler than a similar homogeneous problem. For example, because $\mathcal{L}^{(\mathrm{oi})}$ is nonsingular, there is no problem of compatibility relations for the boundary conditions [38]. Any choice of distribution function on the boundary will generate a unique steady-state solution. The same considerations apply to the evaluation of the transient response of an open system by integrating (26) with respect to t. The solution is unique and, as we have seen, stable.

6 Mathematical Properties of the Irreversible Model

Let us examine in more detail the properties of the time-irreversible open-system model defined by (26) and (29). First, let us note that the Wigner function derived from a steady-state (42) or transient solution of (26) is purely real-valued, because both the Liouville equation (26) and the boundary conditions (29) are purely real. This implies that the

corresponding density matrix is Hermitian, as required.

Now consider the domain upon which the model is defined, as contrasted to the domain of a spatially closed system. This is illustrated in Figure 7. For a closed system of length l (bounded by an infinite potential well) the state of the system would be described by a density matrix defined within the square formed by the long-dashed lines. The coordinate rotation from the Wigner-Weyl transformation (23) implies that the domain of the Wigner function maps onto the rotated square shown by the short-dashed lines in the x, x' plane. The density operator is, in effect, a spatial correlation function. The partitioning of a one-dimensional "universe" into a finite system bounded by two semi-infinite reservoirs partitions the domain of the density operator into regions corresponding to various system-system, system-reservoir, and reservoir-reservoir correlations. The domain of the Wigner function does not coincide with that of the system-system density operator, and the Wigner function domain extends into regions which describe system-reservoir correlations. This may well be a necessary characteristic of an acceptable open-system model.

The use of boundary conditions which are localized in phase space naturally raises the question of possible violations of the uncertainty principle. To address this question, let us first note that q and p are not at all the same as x and $p_x = (i/\hbar)\partial/\partial x$ of the underlying wavefunctions. In fact, q and p are the eigenvalues of commuting superoperators. These superoperators are the anticommutator superoperators $\mathcal{X}_{(+)}$ and $\mathcal{P}_{(+)}$ as defined in equation (8), which in the position-space basis are given by

$$\mathcal{P}_{(+)} = \frac{\hbar}{2i}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right) = \frac{\hbar}{i}\frac{\partial}{\partial r}, \tag{44}$$

$$\mathcal{X}_{(+)} = \tfrac{1}{2}(x + x'). \tag{45}$$

It is easy to show that $\mathcal{X}_{(+)}$ and $\mathcal{P}_{(+)}$ commute, and it is quite obvious from Figure 7 that they should, because they involve orthogonal directions.

How, then, does the uncertainty principle affect the Wigner function? The characteristic property of a Wigner function which violates the uncertainty principle is that it contains some states which have negative occupation probabilities. That is, the corresponding density matrix will have at least some negative eigenvalues. Consider, for example, a

Wigner function $f(q,p) = \pi\delta(q)\delta(p)$, which clearly violates the uncertainty principle. The corresponding density matrix is $\rho(x,x') = \delta(x+x')$. If we operate on any antisymmetric state $\psi_a(x) = -\psi_a(-x)$ with this density matrix, we get $-\psi_a(x)$, so -1 is certainly an eigenvalue of ρ, which is thus not a valid density matrix.

Therefore, to represent an acceptable mixed state, the density operator ρ must be nonnegative-definite. A necessary and sufficient condition for ρ to be nonnegative-definite is to have

$$\langle\psi|\rho|\psi\rangle \geq 0, \tag{46}$$

for all states ψ. The expectation value can be rewritten as an operator inner product (4) by defining the projection operator $P_\psi = |\psi\rangle\langle\psi|$:

$$\langle\psi|\rho|\psi\rangle = \mathrm{Tr}(P_\psi\rho) = \langle P_\psi\|\rho\rangle. \tag{47}$$

Then the condition (46) can be transformed into the Wigner-Weyl representation using (39) to obtain the condition

$$\int dq \int dp\, f(q,p) f_\psi(q,p) \geq 0, \tag{48}$$

where f_ψ is the Wigner function for the pure state ψ, for all ψ. Note that the condition that f is nonnegative definite as an operator does *not* imply that $f(q,p) \geq 0$. It is well known that the Wigner function can take negative values [35], and that such negative values are related to quantum interference.

Now, does the procedure of directly solving for the Wigner function under inhomogeneous boundary conditions lead to a nonegative-definite $f^{(\mathrm{dc})}$ operator? It does not appear that there is a mathematical demonstration which guarantees that such is the case, and thus it is probably possible to define a case of the present open-system model which does violate the uncertainty principle. Let us explore some of the considerations which bear upon the question of the nonnegativity of the Wigner function. First, note that the nonnegativity of $f^{(\mathrm{dc})}$ necessarily involves the nonnegativity of the boundary values, because $f^{(\mathrm{dc})}$ is a linear function of the boundary values as shown by (42). We can speculate that

at least in a semiclassical case $f^{(dc)}$ should be nonnegative-definite if $f^{(l)}_{bdy}$ and $f^{(r)}_{bdy}$ are nonnegative. To establish the plausibility of the idea, let us consider the classical case. The properties of the classical Liouville equation (26) employing the open system boundary conditions (29) are essentially the same as those of the quantum case with respect to the eigenvalue spectrum of the Liouville operator and the stability of the resulting solutions. If we assume that there is no damping within the system, then the classical Liouville theorem holds within the system, and the distribution function f_{cl} is constant along the classical trajectories. Any trajectory passing through a boundary must in fact pass through a boundary twice, once as an incoming particle and once as an outgoing particle (otherwise a density would have to build up in violation of the Liouville theorem). Such trajectories cover the phase space, except for those regions which correspond to any bound orbits. Because f_{cl} is constant along a trajectory and its value is fixed by the boundary condition, f_{cl} must be nonnegative if and only if the boundary values are nonnegative. The values of f_{cl} in regions corresponding to bound states will be nonegative if and only if the initial values of f_{cl} (with respect to time) are nonnegative.

How might these considerations be modified in a quantum-mechanical system? Or, how can one get into trouble applying the open-system boundary conditions to a quantum system? The only obvious case would be if one attempted to apply the boundary conditions (29) in a region where there were strong interference effects, such as standing waves. We can easily imagine that, for example, forcing f to have a large density at a boundary point where a node in the density should occur might introduce spurious states with negative occupation. To avoid such situations, one should apply (29) only in reasonably classical regions of a system; in practice this means at least a thermal de Broglie wavelength away from any abrupt feature of the potential. Note that such considerations imply that the present open system model is appropriate for finite temperatures and that it is likely to break down as the temperature approaches zero.

Now let us examine in more detail the mathematical structure of the model which results from the time-irreversible boundary conditions. The discrete expression for the

drift term $\mathcal{T}$ of the Liouville equation (38) has the form of a master operator, by which I mean an operator which could appear on the right-hand side of a master equation [39] which describes a Markovian process. Such an operator, when applied to a distribution function, has the effect of removing some fraction of the density in each possible state and redistributing that fraction among the other possible states. For a finite, discrete model the properties of the matrix representing a master operator are:

$$\begin{aligned} a_{ii} &\leq 0, \\ a_{ij} &\geq 0 \quad \text{for } i \neq j, \\ \sum_j a_{ij} &\leq 0. \end{aligned} \tag{49}$$

In the last condition the row sum is actually equal to zero except for those states i which can lose density to an external reservoir, as is the case for the open system model on the outflowing boundaries. All the eigenvalues of a matrix satisfying the conditions (49) will have nonpositive real parts. This may be readily demonstrated by appealing to Gerschgorin's theorem [45], which states that every eigenvalue of a matrix A lies in at least one of the circular discs (in the complex plane) with centers at a_{ii} and radii $\sum_{i \neq j} |a_{ij}|$. Because a_{ij} is negative for $i = j$ and positive for $i \neq j$ and is real for all i and j, we find that the real part of each eigenvalue λ_k must satisfy

$$a_{ii} - \sum_{j \neq i} a_{ij} \leq \Re\lambda_k \leq a_{ii} + \sum_{j \neq i} a_{ij} = \sum_j a_{ij} \leq 0, \tag{50}$$

for some i. Thus, $\Re\lambda_k \leq 0$ for all k. The fact that the row sums in $\mathcal{T}$ for the outflow boundaries are less than zero makes $\mathcal{T}$ nonsingular. (In a master operator describing a closed system, all the row sums would be zero, which implies that the determinant would be zero, so there must be an eigenvalue equal to zero.)

The fact that the upwind discretization generates a master operator is the fundamental reason for its success, both in the present context and in the more traditional applications of transport theory [31,32]. Now, in the quantum case, the complete Liouville operator $\mathcal{L}$ (in the Wigner-Weyl representation) cannot be a master operator, because we know that the Wigner distribution can have negative values, which a master operator would not permit.

As we have noted, the quantum interference phenomena enter the Wigner distribution via the potential superoperator $\mathcal{V}$. The fundamental result of the present work is the demonstration in Fig. 6 and equations (40)-(41) that the Markovian model which follows from the irreversible boundary conditions (29) introduces the necessary stability properties in the quantum case as well as in the much more obvious classical case.

It is interesting to consider the form $\mathcal{T}$ assumes upon transformation back to a real-space density matrix representation. For this purpose let us assume that we have defined the Wigner function on a discrete basis with respect to q and on a continuum basis with respect to p. Then $\mathcal{T}$ is given by

$$(\mathcal{T}f)(q,p) = -\frac{p}{m\Delta_q} \times \begin{cases} f(q+\Delta_q,p) - f(q,p) & \text{for } p<0 \\ f(q,p) - f(q-\Delta_q,p) & \text{for } p>0 \end{cases} . \tag{51}$$

To transform this back to the density matrix representation, we must evaluate

$$(\mathcal{T}\rho)(q,r) = \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar}\, e^{ipr/\hbar}\, (\mathcal{T}f)(q,p), \tag{52}$$

with (25) substituted for f. [To simplify the resulting expressions, we will express the arguments of ρ in terms of q and r of equation (23) and Figure 7.] Evaluation of (52) requires the formula

$$\int_0^{\infty} \frac{dp}{2\pi\hbar}\, p\, e^{ipr/\hbar} = \hbar\frac{\partial}{\partial r}\left[\frac{1}{2\pi}\wp\frac{1}{r} - \frac{i}{2}\delta(r)\right] \tag{53}$$

and its complex conjugate. Letting Δ_q approach zero we find

$$(\mathcal{T}\rho)(q,r) = \frac{\hbar}{m}\frac{\partial}{\partial r}\left[i\frac{\partial\rho(q,r)}{\partial q} + \frac{\Delta_q}{2\pi}\wp\int_{-\infty}^{\infty}\frac{dr'}{r-r'}\frac{\partial^2\rho(q,r')}{\partial {q'}^2}\right]. \tag{54}$$

The second term in (54) contributes an anti-Hermitian component to $\mathcal{L}$. The appearance of $\partial^2/\partial q^2$ in this term is reminiscent of the "numerical viscosity" which is a property of some finite-difference formulations of transport equations [44]. The principal-value integral in (54) has the desired effect of distinguishing the sign of the momentum of the states present in ρ. To see this, suppose that there is a term $|k\rangle\langle k| = e^{ikr}$ contained in ρ. One could evaluate its contribution to the integral in (54) by contour integration, closing the contour in the upper or lower half-plane if k were positive or negative, respectively. But then

the sign of the contribution of the pole on the real axis changes as the sign of k changes. The anti-Hermitian term vanishes, except possibly for a surface contribution, in the limit $\Delta_q = 0$.

This description of open systems in terms of $\rho(x, x')$ has not yet been developed into a workable model. However, there is a strong motivation for doing so in the context of semiconductor heterostructures (which provided the original motivation for this work). In such a structure the electron energy-momentum relation can be considerably more complex than a simple parabola, and it changes from one material to another in ways which cannot be represented by a shift in the local potential. The simplest example of such an effect is the change in effective mass as an electron crosses a heterojunction. This may be modeled by writing the kinetic-energy term of the Hamiltonian as

$$T = -\frac{\hbar^2}{2}\frac{\partial}{\partial x}\frac{1}{m^*(x)}\frac{\partial}{\partial x},$$

which thus locally modifies the Liouville operator for $\rho(x, x')$, but which produces nonlocal terms in the Liouville operator for the Wigner function [40]. More complex features of the energy-band structure can be modeled by any of a number of localized-basis-function schemes which may require more than one basis function per lattice site. Such schemes could easily fit into an approach expressed in terms of $\rho(x, x')$, but it is not at all obvious how to incorporate such effects into the Wigner function.

Of more general interest is the appearance of (53) in the deductive chain leading to (54). Such a relation, more often expressed in the form

$$\frac{1}{\omega \pm \epsilon} = \wp\frac{1}{\omega} \mp i\pi\delta(\omega), \tag{55}$$

is usually encountered in the analysis of irreversible quantum phenomena. It is the mathematical expression of the fact that a continuum of states (and therefore of frequencies) provides enough degrees of freedom that a Poincaré recurrence can be postponed indefinitely. It appears in the analysis of behavior in the time and frequency domains, and is used to express the initial conditions which lead to irreversible behavior: no advanced waves in electrodynamics [41], or adiabatic switching-on in many-body theory [42,43]. In

the present model such a relation appears in the position and momentum domains and expresses the effects of the spatial boundary conditions.

7 Superoperator Symmetry and Conserved Quantities

One of the benefits of the time-irreversible open-system boundary conditions is that they provide an alternative to the use of periodic boundary conditions in the analysis of quantum-transport phenomena. The great disadvantage of periodic boundary conditions is that they do not address the case in which the potential varies significantly across a system. That is, their use restricts one to the study of low-field phenomena. It has been pointed out [46] that quasi-periodic boundary conditions (*i.e.*, periodic within a phase factor which can be removed by a gauge transformation) are necessary if the momentum operator is to be Hermitian on a finite domain. The present work demonstrates that far-from-equilibrium phenomena can be modeled by employing a non-Hermitian momentum superoperator.

The connection between symmetries and conservation laws is undoubtedly one of the great conceptual triumphs of the quantum theory. However, if one is faced with the task of describing the behavior of a nonconservative system, the inability to modify or violate the conservation laws becomes an obstacle to defining a realistic model, rather than a benefit. The problem is that one wants a model whose solutions stably approach a steady state, which requires complex-valued eigenvalues, but the expectation values of physical observables should be real. The present analysis of open-system models demonstrates that these conflicting requirements can be accommodated at the kinetic level, because the roles of generating the dynamical evolution and evaluating observables are filled by different *superoperators*. If we reexamine the models described above, we find that the dynamic effects such as generating time evolution or moving density by current flow are described by commutator superoperators, and these are the superoperators which become non-Hermitian when one incorporates interactions with the outside world. The measurement of the expectation values of observables is done by anticommutator superoperators, and these, ideally, remain Hermitian.

The general principles that commutators generate motion and can be non-Hermitian and that anticommutators measure observables can be seen in two specific cases: the superoperators generated by the momentum operator and by the Hamiltonian. We have already observed that the kinetic energy term of the Liouville equation (3) can be factored into a sum and difference of derivatives (14). These are just the commutator and anticommutator superoperators $\mathcal{P}_{(-)}$ and $\mathcal{P}_{(+)}$. The superoperator $\mathcal{P}_{(+)}$ was defined in equation (44). It will be Hermitian if we restrict our attention to density matrices whose off-diagonal elements approach zero for large $x - x'$. Such density matrices describe normal systems (as opposed to superconducting, or with some other long-range coherent effect) at nonzero temperature. In such normal cases $\mathcal{P}_{(+)}$ produces the real-valued factor p in the drift term (31). The commutator,

$$\mathcal{P}_{(-)} = \frac{\hbar}{i}\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial x'}\right) = \frac{\hbar}{i}\frac{\partial}{\partial q}, \tag{56}$$

generates the gradient in $\mathcal{T}$ and is thus the superoperator which is rendered non-Hermitian by the boundary conditions (29).

We can also see the dichotomy of function between $\mathcal{P}_{(+)}$ and $\mathcal{P}_{(-)}$ by examining the elementary quantum continuity equation, which is conventionally written

$$\frac{\partial}{\partial t}\psi^*\psi = -\frac{\partial}{\partial x}(\psi^* J \psi), \tag{57}$$

where

$$(\psi^* J \psi) = \frac{\hbar}{2mi}\left(\psi^*\frac{\partial \psi}{\partial x} - \frac{\partial \psi^*}{\partial x}\psi\right). \tag{58}$$

By now we should readily recognize the presence of $\mathcal{P}_{(+)}$ in the current density operator J. In fact, the current density is much more naturally regarded as a superoperator, $\mathcal{J} = \mathcal{P}_{(+)}/m$, and we see $\mathcal{P}_{(+)}$ in the role of measuring an observable. At the kinetic level the continuity equation is linear in terms of the density matrix ρ and is simply the Liouville equation evaluated along the diagonal $x = x'$. To see this, we rewrite the Liouville equation for ρ (3), using the factorization of the kinetic energy term (14) and the definitions of $\mathcal{J}$ and $\mathcal{P}_{(-)}$, as

$$\frac{\partial \rho(x,x')}{\partial t} = -\left[\frac{\partial}{\partial[\frac{1}{2}(x+x')]}\mathcal{J}\rho\right](x,x') + \frac{1}{i\hbar}[v(x) - v(x')]. \tag{59}$$

If we denote the evaluation of an operator kernel for $x = x'$ by angle brackets, $\langle\rho\rangle_x = \rho(x,x)$, and apply this operation to the Liouville equation, we obtain

$$\frac{\partial\langle\rho\rangle_x}{\partial t} = -\frac{\partial\langle\mathcal{J}\rho\rangle_x}{\partial x}, \tag{60}$$

which is a familiar form of the continuity equation.

The balance equations derived from higher moments of the Liouville equation are obtained by operating on the equation with $\mathcal{P}_{(+)}$ (or $\mathcal{J}$) and evaluating the resulting expression along the diagonal. This is equivalent to the phase-space procedure of multiplying by some power of p and then integrating over all p. The first moment equation is thus

$$m\frac{\partial\langle\mathcal{J}\rho\rangle_x}{\partial t} = -\frac{\partial}{\partial x}\langle\Pi\rho\rangle_x - \frac{\partial v}{\partial x}\langle\rho\rangle_x, \tag{61}$$

where $\Pi = \mathcal{P}_{(+)}^2/m$ is the momentum flux density. (For two- or three-dimensional models, the direct product of the two factors of $\mathcal{P}_{(+)}$ is taken and Π will be a tensor [47].) Equation (61) is identical to its classical counterpart. If we integrate it with respect to x (over the domain $0 < x < l$), we obtain a generalization of Ehrenfest's theorem to the case of an open system:

$$m\frac{\partial}{\partial t}\int_0^l \langle\mathcal{J}\rho\rangle_x\, dx = -\int_0^l \frac{\partial v}{\partial x}\langle\rho\rangle_x\, dx + \langle\Pi\rho\rangle_0 - \langle\Pi\rho\rangle_l\,. \tag{62}$$

The last two terms represent the effect of opening the system: a flux of momentum density through the boundaries of the system will affect the current flow within the system.

The same dichotomy between commutator and anticommutator superoperators can be seen in the case of the superoperators generated by the Hamiltonian H. Of course $\mathcal{H}_{(-)}$ is just the Liouville superoperator $\mathcal{L}$, and we have examined at length the need for a departure from Hermiticity in the case of $\mathcal{L}$. We have not yet encountered a need for the anticommutator $\mathcal{H}_{(+)}$. One place it does occur is in a generalization of the Bloch equation (10) to the case of an open system. If one attempts to compute an equilibrium density matrix as a finite segment of a much larger system by modifying the boundary conditions on ρ in the Bloch equation, one quickly discovers that product $H\rho$ must be symmetrized to obtain sensible answers. Thus, the Bloch equation becomes

$$\partial\rho_{\rm eq}/\partial\beta = -\tfrac{1}{2}(H\rho_{\rm eq} + \rho_{\rm eq}H) = -\mathcal{H}_{(+)}\rho_{\rm eq}. \tag{63}$$

If the time-reversible open-system boundary conditions (12) are applied to the Bloch equation, one obtains a quite useful method for evaluating the equilibrium density matrix (in contrast to the disastrous effect these boundary conditions have upon the time evolution). Taking into account our particle-density normalization of ρ, the correct Bloch equation is

$$\partial\rho_{\rm eq}/\partial\beta = -(\mathcal{H}_{(+)} - \mu)\rho_{\rm eq}, \tag{64}$$

with the initial condition

$$\rho_{\rm eq}|_{\beta=0} = \delta(x - x'). \tag{65}$$

When this equation is integrated, the resulting densities in regions of constant potential are found to be equal to the semiclassically expected value $(m/2\pi\hbar^2\beta)^{1/2}\exp[\beta(\mu - v)]$. An example of an equilibrium density matrix obtained from such a calculation is illustrated in Figure 8.

Here we see that again the anticommutator superoperator appears in the role of evaluating an observable, in this case for the purpose of evaluating the energy and thus the occupation probability of the possible states. We might expect that, for this purpose, $\mathcal{H}_{(+)}$ ought to be Hermitian. Unfortunately, this is not the case when the boundary conditions (12) are applied. While the eigenvalues of $\mathcal{H}_{(+)}$ are clustered near the real axis, they are, in general, complex. However, this does not seem to produce any unphysical effects. The results of calculations employing (64) appear to be quite reasonable, and all the cases which I have examined can be understood in terms of the behavior of the underlying wavefunctions, as in the case shown in Figure 8.

Having noted that $\mathcal{H}_{(+)}$ appears in the evaluation of the equilibrium density matrix, we can address a point raised by Dahl [48]. It is that $\mathcal{L}$, by itself, does not define a unique eigenvalue problem for a quantum system; but together with $\mathcal{H}_{(+)}$, it does define such a problem. This consideration enters the present problem only for bound states localized within the open system [30]. As noted earlier, such states would lead to a nontrivial null space of $\mathcal{L}$. The occupation of such states would have to be determined as an initial condition, such as an equilibrium distribution evaluated using $\mathcal{H}_{(+)}$.

8 Design of Discrete Numerical Models

The present work employs numerical computation and modeling for a purpose for which it is not often employed: as the primary mode of investigating the structure and consequences of a physical theory. The more traditional mode of investigation is, of course, to maximize the use of analytical mathematics and resort to numerical techniques only when the opportunities for analysis are exhausted, or when it is necessary to evaluate those complicated expressions which express an analytical solution. Any approach to describing physical phenomena will be successful only for some subset of those phenomena and will be otherwise ineffective. Because analytical mathematics is such a widely used tool, its domain of success has been extensively explored; this domain, of course, consists of those problems with sufficient symmetry to admit analytic solutions and those problems which can be regarded as small perturbations on analytically soluble problems. For statistical phenomena this generally means thermal equilibrium of analytically tractable systems and its near neighborhood. Numerical simulation techniques which are nonperturbative are better able to address more complex structures and/or far-from-equilibrium states. Because the study of discrete numerical models is not widely practiced, it is worth examining the principles by which such models may be constructed, using the present open-system model as an example.

It is customary to regard discrete numerical models, such as finite-difference models for partial differential equations, as approximations to the "truth" embodied in the continuum formulation of the problem [49]. Such a discrete model can represent the continuum solution only to within an accuracy which is proportional to some power of the mesh spacing (or other appropriate measure of the coarseness of the discrete model). This tends to lead one to believe that the physics of the situation can be represented only to a given order of accuracy, so that such expressions as conservation laws (or balance equations) will be satisfied only to that order (see, for example, [37]). A corollary to this view is that higher-order approximations produce better models. Such is often not the case [44].

In fact, discrete models can be constructed so as to exactly satisfy most of the balance

equations and other global conditions which govern the behavior of the real physical system. An example of a "global condition" which is not a balance equation or conservation law is the stability requirement that there be no eigenvalues of $\mathcal{L}$ with positive imaginary parts. It is absolutely essential that a discrete model exactly satisfy this condition. On the other hand, a model may be adequate for many purposes even if it only satisfies the momentum balance equation to the order of some power of the mesh spacing. Thus, one can articulate a general approach to the design of discrete models: First, identify all the relevant relations which the model should satisfy; then, establish the relative importance of these relations; finally, design a discretization scheme which satisfies the most important relations and trade off the less important relations against the computational resources required. The problem with this prescription is that there is no reliable way to ascertain that one has identified *all* the relevant relations or has prioritized them correctly. Nevertheless, the articulation of such a general approach is of value in a field which is often described as being "an art as much as a science" [44].

To observe how a discrete model can be made to exactly satisfy conservation laws and other global conditions, let us again consider the discrete Liouville equation with irreversible boundary conditions defined by equation (43). As we have noted, the total number of particles is not conserved because the system is open. However, particles can be lost or gained only by current flow through the boundaries, so particle conservation is described *locally* by the continuity equation, which is obtained from the Liouville equation for the Wigner function (26) by integrating over the momentum p. From the normalization convention for the density matrix and the definition of the Wigner function (25), we find that the expressions for the density of particles $n(q)$ and the current density $J(q)$ at position q are

$$n(q) = \rho(q,q) = \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} f(q,p), \tag{66}$$

$$J(q) = \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} \frac{p}{m} f(q,p). \tag{67}$$

In integrating the Liouville equation over p, the contribution of the potential term is zero

because of the antisymmetry of its kernel. We thus obtain the continuity equation

$$\partial n/\partial t = -\partial J/\partial q. \tag{68}$$

To obtain the continuity equation for the discrete model we follow the same steps. The density is obtained by summing f over the discrete values p_k:

$$n_j \equiv n(q_j) = \frac{\Delta_p}{2\pi\hbar}\sum_k f_{jk}. \tag{69}$$

In a discrete model the current density is most naturally regarded as a quantity which is defined on each *interval* between adjacent mesh points, rather than on the mesh points themselves. Thus, the divergence of the current density is a difference taken between adjacent intervals and is associated with their common mesh point. Let us denote the current on the interval between q_j and q_{j+1} by $J_{j+1/2}$. Then we must define $J_{j+1/2}$ to be

$$J_{j+1/2} = \frac{\Delta_p}{2\pi\hbar}\left(\sum_{k|p_k<0}\frac{p_k}{m}f_{j+1,k} + \sum_{k|p_k>0}\frac{p_k}{m}f_{j,k}\right). \tag{70}$$

The moment of the Liouville equation becomes

$$\frac{\partial n_j}{\partial t} = -\frac{1}{\Delta_q}\left(J_{j+1/2} - J_{j-1/2}\right) - \frac{\Delta_p}{2\pi\hbar}\sum_{k=1}^{N_p}\sum_{k'=1}^{N_p} V_{j;k,k'}f_{q,k'}.$$

To show that the contribution from the potential operator V vanishes, let us consider the sum over k first. The sum can be reordered and then $V_{j,k}$ can be expanded using equation (33):

$$\sum_{k=1}^{N_p} V_{j;k,k'} = \sum_{k=1}^{N_p} V_{j,k} = \frac{2}{N_p}\sum_{j'=1}^{N_q/2}\sum_{k=1}^{N_p}\sin\left(\frac{2k\Delta_p j'\Delta_q}{\hbar}\right)(v_{j+j'} - v_{j-j'}).$$

Now, this sum will vanish if the sum

$$\sum_{k=1}^{N_p}\sin\left(\frac{2k\Delta_p j'\Delta_q}{\hbar}\right)$$

vanishes, which happens if $(2N_p\Delta_p\Delta_q)/\hbar = 2\pi$, and Δ_p was defined so as to satisfy this relation. This is the Fourier completeness relation mentioned earlier. Thus, the discrete model exactly satisfies the continuity equation

$$\frac{\partial n_j}{\partial t} = -\frac{1}{\Delta_q}\left(J_{j+1/2} - J_{j-1/2}\right). \tag{71}$$

The only limit on the precision of this relationship is the arithmetic roundoff error, which is generally several orders of magnitude smaller than typical discretization errors.

One begins to encounter the limits of a simple discrete model when the momentum balance (first moment) equation (61) is considered. To evaluate the rate of change of current density, insert the discrete Liouville equation (43) into the definition of $J_{j+1/2}$ (70). One then obtains

$$\begin{aligned} m\frac{\partial J_{j+1/2}}{\partial t} = & - \frac{1}{\Delta_q}(\Pi_{j+1} - \Pi_j) \\ & - \frac{\Delta_p}{2\pi\hbar^2}\left(\sum_{k|p_k<0} p_k \sum_{k'} V_{j+1;k,k'} f_{j+1,k'} + \sum_{k|p_k>0} p_k \sum_{k'} V_{j;k,k'} f_{j,k'}\right), \end{aligned} \tag{72}$$

where

$$\Pi_j = \frac{\Delta_p}{2\pi\hbar}\left(\sum_{k|p_k<0} \frac{p_k^2}{m} f_{j+1,k} + \sum_{k|p_k>0} \frac{p_k^2}{m} f_{j-1,k}\right). \tag{73}$$

Note that the requirements of consistency in the discretization scheme imply that Π_j, which one might expect to depend only upon the values of f at q_j, actually depends upon the values of f at q_{j-1} and q_{j+1}. This sort of spreading over the domain becomes worse as higher moments are considered. Now consider the potential terms in (72). For simplicity, let us neglect the different j indices required by the form of J and simply evaluate

$$\frac{\Delta_p}{2\pi\hbar^2}\sum_{k=1}^{N_p} p_k \sum_{k'=1}^{N_p} V_{j;k,k'} f_{j,k'} = \left[\sum_{j'=1}^{N_q/2} \frac{\pi}{2N_p\Delta_q} \cot\left(\frac{\pi j'}{N_p}\right)(v_{j+j'} - v_{j-j'})\right]\left[\frac{\Delta_p}{2\pi\hbar}\sum_{k'=1} N_p f_{j,k'}\right], \tag{74}$$

where equation (33) is again used and the sums reordered as before. Now, in the continuum case (61) this expression reduces to $(\partial v/\partial q)n$. The discrete expression (74) shows a functional of v (the first bracketed factor) times n. If we consider only the first term of the sum over j' and take $\cot\alpha \approx 1/\alpha$ for small α, we get $(v_{j+1} - v_{j-1})/2\Delta_q$, which is just the centered-difference approximation to $\partial v/\partial q$. However, the other terms of the sum are not negligible. While $\pi j'/N_p$ is small, the higher terms just add in more remote approximations to $\partial v/\partial q$. Of course, $\cot\alpha$ approaches zero much more rapidly than $1/\alpha$ as α approaches $\pi/2$. Thus, there is a natural cutoff of these higher terms so long as $j' \lesssim N_p/2$. This helps to explain the significance of the limit of the j' summation of (33). The value

of $N_q/2$ was originally chosen for the upper limit of this sum on the purely empirical basis that the results were most credible with this value, and multiples of N_q were investigated because the summation is carried out in position space. However, most calculations have taken $N_p \approx N_q$, so these conditions are approximately equivalent. However, the significant result is that the momentum balance equation (61) is not satisfied exactly by the discrete model.

Defining discrete models in such a way as to exactly satisfy the appropriate physical laws leads to practical benefits in the evaluation of such models. Such formulations are very "robust" in the sense that they do not catastrophically fail when applied to cases which are in some way severe. A particularly useful case is the model which has a very few nodes. We have seen how such models can be quite useful in investigating the structure of a theory. While there may not be enough degrees of freedom to accurately represent the behavior of a realistic system, the qualitative behavior of such a model can contain useful information. Such models require very modest computational resources, of course.

9 Conclusions

The central conclusion of the present work is that an open system, in the sense of one which exchanges particles with its environment through spatially localizable interfaces, is necessarily irreversible. The reasoning behind this conclusion is a *reductio ad absurdum* argument. We have seen that a particular reversible model of an open system possesses unphysical instabilities. The mathematical properties underlying these instabilities, the existence of complex eigenvalues of non-Hermitian superoperators and the requirement that these occur in conjugate pairs due to time-reversal symmetry, are sufficiently general that we should expect such instabilities in any reversible model. Thus, physically acceptable models of open systems must be inherently time-irreversible.

A particular irreversible open-system model was presented, and its stability was demonstrated. The irreversibility of this model follows from making a distinction between particles entering and leaving the system. Similar ideas, generally applied in the

time domain, are the basis for the established theories of irreversibility and dissipation. The present work demonstrates that spatial boundary conditions can be used to introduce irreversibility in a way which is very similar to that by which temporal initial conditions do so.

The present study of the kinetic theory of open systems helps to clarify the roles of superoperators generated by the commutator and anticommutator of a physical observable. It was demonstrated that, at the kinetic level, only the commutator superoperators should acquire non-Hermitian parts to model irreversible phenomena. Anticommutator superoperators ideally remain Hermitian and are used to evaluate expectation values.

This work is certainly not an exhaustive examination of the theory of open systems. Undoubtedly, many more approaches to the subject can be formulated. However, one should note that the significant behaviors of an open system involve a strong coupling between the system and its environment and large deviations from equilibrium within the system. It thus appears unlikely that perturbative approaches will contribute very much to the theory of such systems. Thus, numerical models will have to be the mainstay of such investigations.

References

[1] R. C. Tolman, *The Principles of Statistical Mechanics* (Dover, New York, 1979), sec. 140.

[2] G. V. Chester, Rep. Prog. Phys. **26**, 411 (1963).

[3] I. Oppenheim, K. E. Shuler, and G. H. Weiss, *Stochastic Processes in Chemical Physics: The Master Equation* (MIT Press, Cambridge, Mass., 1977) and reprints included therein.

[4] E. B. Davies, *Quantum Theory of Open Systems* (Academic Press, London, 1976), and references therein.

[5] W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973).

[6] H. Haken, "Cooperative Phenomena in Systems Far from Thermal Equilibrium and in Nonphysical Systems," Rev. Mod. Phys. **47**, 67 (1975).

[7] T. S. Reynolds, *Stronger than a Hundred Men, A History of the Vertical Water Wheel* (Johns Hopkins Univ. Press, Baltimore, 1983), ch. 4.

[8] H. W. Dickinson, *A Short History of the Steam Engine* (Cambridge Univ. Press, Cambridge, 1938), ch. 6.

[9] See, for example, P. Horowitz and W. Hill, *The Art of Electronics* (Cambridge Univ. Press, Cambridge, 1980).

[10] *Webster's New Collegiate Dictionary* (G. & C. Merriam, Springfield, Mass., 1981).

[11] E. A. Marland, *Early Electrical Communication* (Abelard-Schuman Ltd., London, 1964).

[12] P. J. Roache, *Computational Fluid Dynamics* (Hermosa Publishers, Albuquerque, NM, 1976), ch. III, sec. C.

[13] A. G. Milnes and D. L. Feucht, *Heterojunctions and Metal-Semiconductor Junctions* (Academic Press, New York, 1972), sec. 9.9.

[14] J. Bardeen, "On the Theory of the A. C. Impedance of a Contact Rectifier," Bell Syst. Tech. J. **28**, 428 (1949).

[15] W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," Bell Syst. Tech. J. **28**, 435 (1949).

[16] S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer-Verlag, Wien, 1984), sec. 5.1.

[17] E. L. Wolf, *Principles of Electron Tunneling Spectroscopy* (Oxford Univ. Press, New York, 1985).

[18] R. Mains, private communication.

[19] L. E. Reichl, *A Modern Course in Statistical Physics* (University of Texas Press, Austin, 1980).

[20] R. Zwanzig, "On the Identity of Three Generalized Master Equations," *Physica* **30**, 1109 (1964).

[21] T. M. Apostol, *Calculus*, Vol. II, Multi-Variable Calculus and Linear Algebra (Blaisdell Publishing, Waltham, MA, 1969), ch. 1.

[22] W. R. Frensley, "Simulation of Resonant-Tunneling Heterostructure Devices," J. Vac. Sci. Technol. B **3**, 1261 (1985).

[23] R. P. Feynman, *Statistical Mechanics, A Set of Lectures* (W. A. Benjamin, Reading, Mass., 1972), ch. 2.

[24] A. Messiah, *Quantum Mechanics* (Wiley, New York, 1962) Vol. 1, p. 120.

[25] E M. Conwell, *High Field Transport in Semiconductors* (Academic Press, New York, 1967).

[26] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II, Nonequilibrium Statistical Physics* (Springer-Verlag, Berlin, 1985), ch. 2.

[27] A. O. Caldeira and A. J. Leggett, Physica **A121**, 587 (1983).

[28] I. Langmuir and K. T. Compton, Rev. Mod. Phys. **2**, 123 (1931).

[29] E. Wigner, "On the Quantum Correction For Thermodynamic Equilibrium," Phys. Rev. **40**, 749 (1932).

[30] P. Carruthers and F. Zachariasen, Rev. Mod. Phys. **55**, 245 (1983).

[31] P. J. Roache, *op. cit.*, pp 4–5.

[32] J. J. Duderstadt and W. R. Martin, *Transport Theory* (Wiley, New York, 1979) sec. 8.1.2.

[33] W. R. Frensley, "Wigner-Function Model of a Resonant-Tunneling Semiconductor Device," Phys. Rev. **B36**, 1570 (1987).

[34] W. R. Frensley, "Quantum Transport Calculation of the Small-Signal Response of a Resonant Tunneling Diode," Appl. Phys. Lett. **51**, 448 (1987).

[35] E. P. Wigner, "Quantum-Mechanical Distribution Functions Revisited," in *Perspectives in Quantum Theory*, ed. by W. Yourgrau and A. van der Merwe (MIT Press, Cambridge, Mass., 1971), p.25.

[36] M. Dresden, Rev. Mod. Phys. **33**, 265 (1961).

[37] J. P. Aubert, J. C. Vaissiere, and J. P. Nougier, "Matrix Determination of the Stationary Solution of the Boltzmann Equation for Hot Carriers in Semiconductors," J. Appl. Phys. **56**, 1128 (1984).

[38] C. Lanczos, *Linear Differential Operators* (D. van Nostrand Co., London, 1961), sec. 3.6, sec. 3.8, and sec. 4.16.

[39] L. E. Reichl, *op. cit.*, sec. 6B.

[40] J. R. Barker, D. W. Lowe and S. Murray, in *The Physics of Submicron Structures*, ed. by H. L. Grubin, K. Hess, and D. K. Ferry (Plenum Press, New York, 1984), pp. 277-286.

[41] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964), sec 6.3.

[42] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill Inc., New York, 1971), pp. 59-61.

[43] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957).

[44] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes, The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, 1986) sec. 17.1.

[45] J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford Univ. Press, London, 1965), ch. 2, sec. 13.

[46] D. R. Yennie, "Integral Quantum Hall Effect for Nonspecialists," Rev. Mod. Phys. **59**, 781 (1987), footnote 11 acknowledging private discussion with M. Weinstein.

[47] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, trans. by J. B. Sykes and W. H. Reid (Pergamon Press, London, 1959), ch. 1.

[48] J. P. Dahl, "Dynamical Equations for the Wigner Functions," Technical University of Denmark preprint (1981).

[49] L. Lapidus and G. F. Pinder, *Numerical Solution of Partial Differential Equations in Science and Engineering* (Wiley, New York, 1982), ch. 2.

Figure Captions

Figure 1. Potential used in evaluating eigenvalue spectra of Liouville superoperators in the discrete model. This potential was chosen to have both a driving field and a barrier.

Figure 2. Eigenvalue spectra of the Liouville operator for a small model closed system with the potential shown in Fig. 1. If the system is taken to be conservative, the resulting eigenvalue spectrum is shown in (a). All eigenvalues are purely real, as expected. In (b) a damping term has been added, leading to negative imaginary parts for most eigenvalues.

Figure 3. Eigenvalue spectra for open systems using the boundary conditions of equation (12). If the boundary conditions are changed so as to open the system, nonzero imaginary parts are generated, as in (a). Because the boundary conditions are time-reversible, these imaginary parts occur in conjugate pairs. If a damping term is added as in (b), most, but not all the imaginary parts are negative. The few eigenvalues with positive imaginary parts are sufficient to render the model unstable. Stability can be achieved by increasing the damping rate, leading to the spectrum (c).

Figure 4. Possible boundary conditions for the Liouville equation (26) in phase space. The points at which the boundary values are specified (indicated by a heavy line) can be at $q = 0$ as in (a), at $q = l$ as in (b), or divided between the two boundaries depending upon the sign of p, as shown in (c) and (d). The boundary conditions (c) are, in fact, the appropriate ones for an open system.

Figure 5. Discretization scheme for the kinetic-energy superoperator (drift term) $\mathcal{T}$ in the Wigner representation. The flow of probability between mesh points is indicated by the arrows, which also define the sense of the finite-difference approximation for the gradient. A flow toward the right requires a left-hand difference and *vice versa*. This is the "upwind" difference scheme and is uniquely determined by the form of the boundary conditions (29).

Figure 6. Eigenvalue spectrum for a model open system with irreversible boundary conditions. All eigenvalues have negative imaginary parts, verifying that the model is stable, despite the fact that no damping is yet included.

Figure 7. Domain of the density matrix and the Wigner distribution function. The arguments of the density matrix are x and x'. The Wigner function is obtained by transforming to the coordinates q and r, followed by a Fourier transform with respect to r. The long-dashed lines indicate the system-reservoir boundaries, and they partition the domain into regions corresponding to the various system-system, system-reservoir and reservoir-reservoir correlations. The short-dashed lines represent the boundaries of the domain of the Wigner-distribution-function model. Note that the Wigner function includes contributions from regions which represent correlations with the reservoirs.

Figure 8. Equilibrium density matrix obtained by numerically integrating the generalized Bloch equation (64) subject to the reversible open-system boundary conditions (52). The potential, displayed above, represents the sort of features which are now realizable using semiconductor heterostructure technology. The chemical potential μ is indicated by the dashed line. The calculation employed parameters appropriate for the $Al_xGa_{1-x}As$ system at 77 K. The three energy barriers create two identical "quantum wells," bounded by contacting layers. The lowest energy states in these wells are pushed toward higher energy by size quantization, which reduces the electron density in the wells via the Boltzmann factor. The shallow peaks off the diagonal measure the correlation between the phase the electron at different positions, and indicate in the present case that the symmetric combination of the well states has a greater occupation factor than the antisymmetric combination.

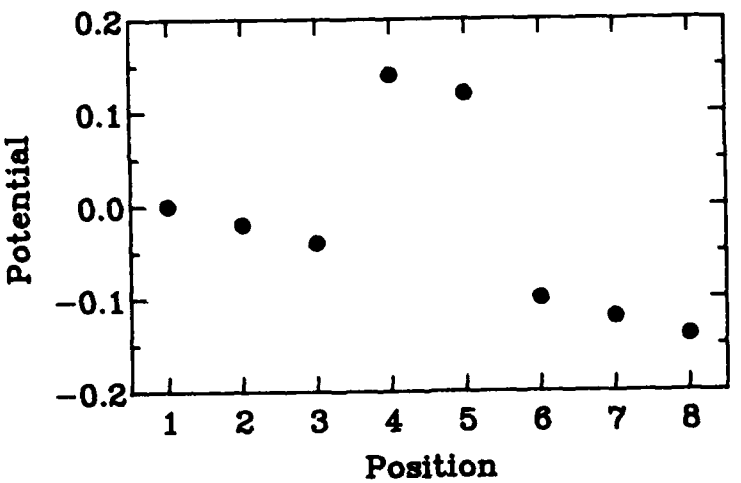


Figure 1

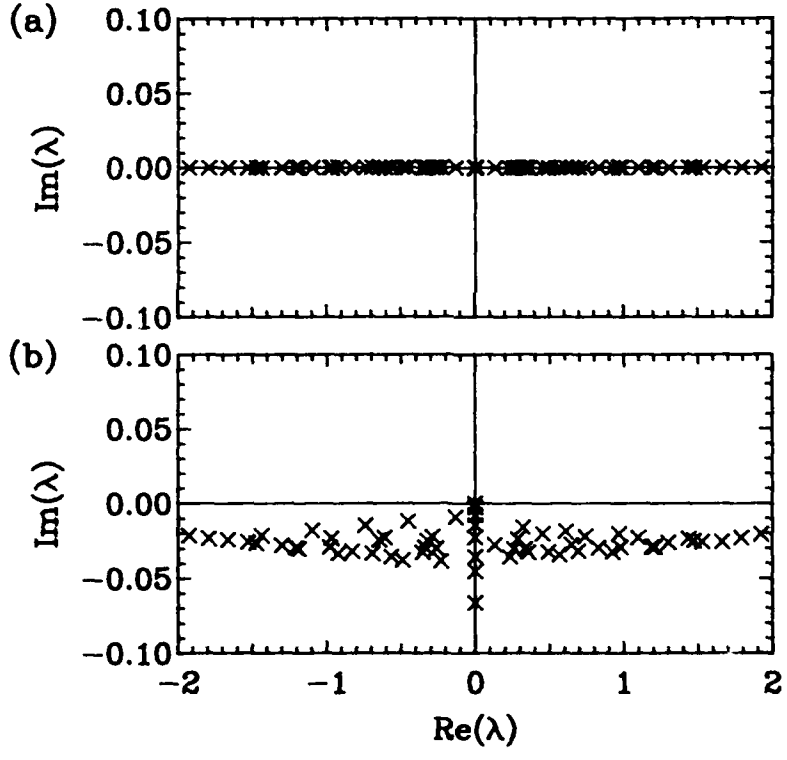


Figure 2

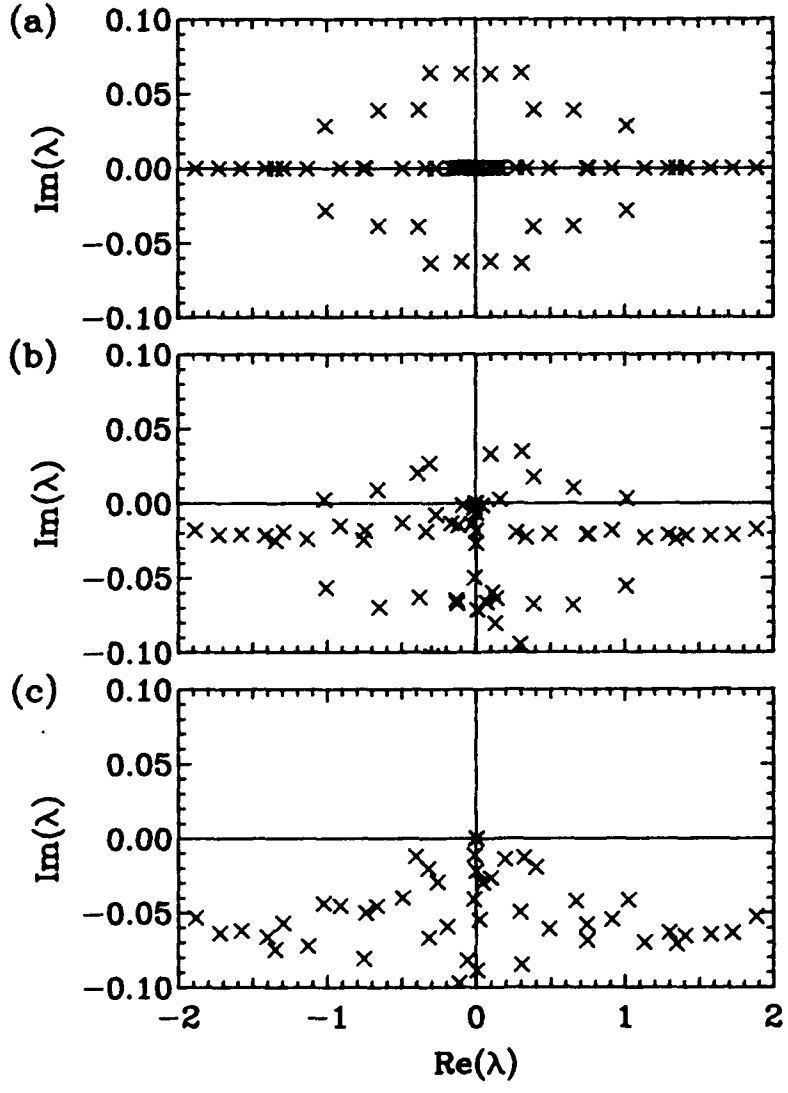


Figure 3

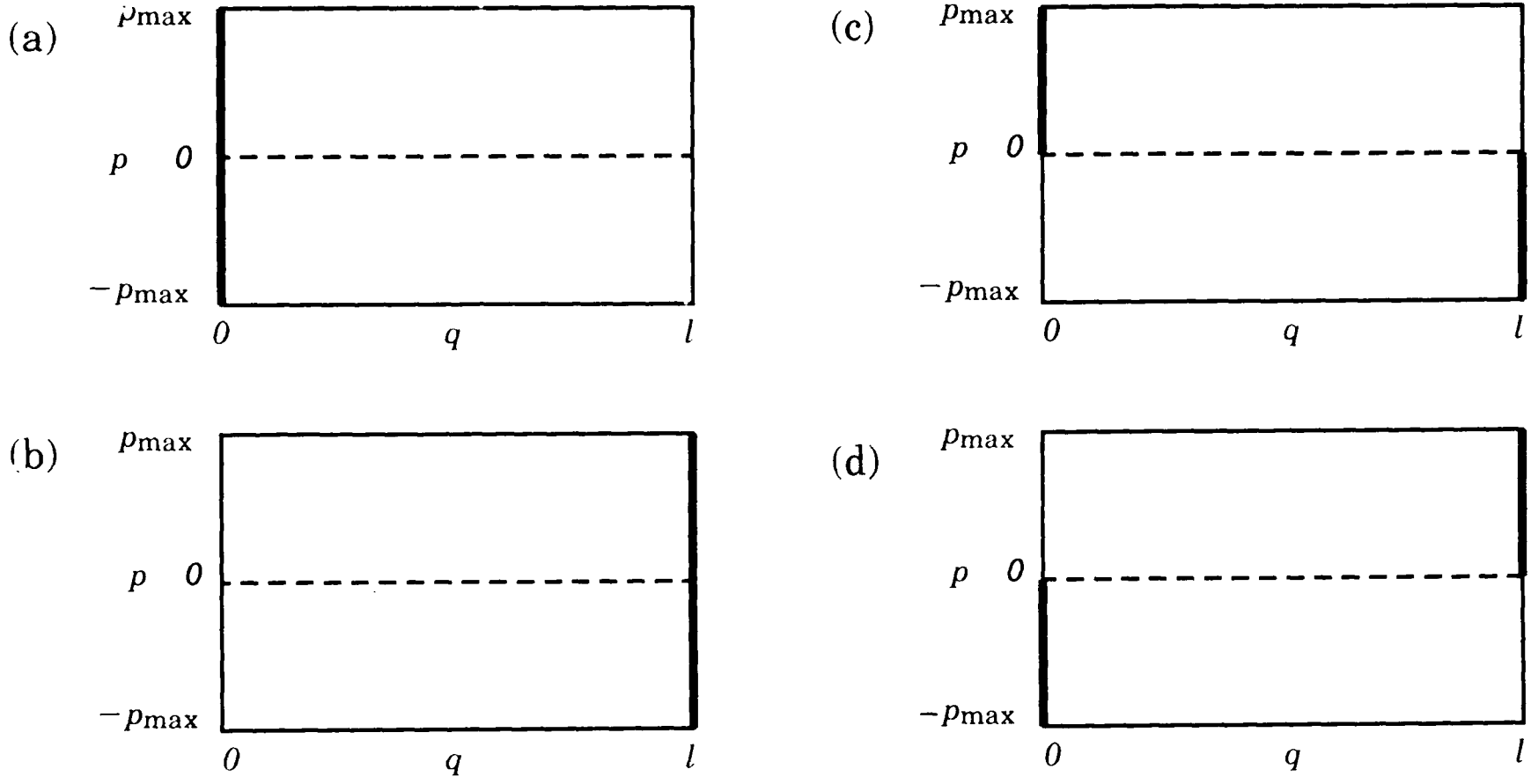

(a)
pmax
p 0
−pmax
0
q
l
(c)
pmax
p 0
−pmax
0
q
l
(b)
pmax
p 0
−pmax
0
q
l
(d)
pmax
p 0
−pmax
0
q
l

Figure 4

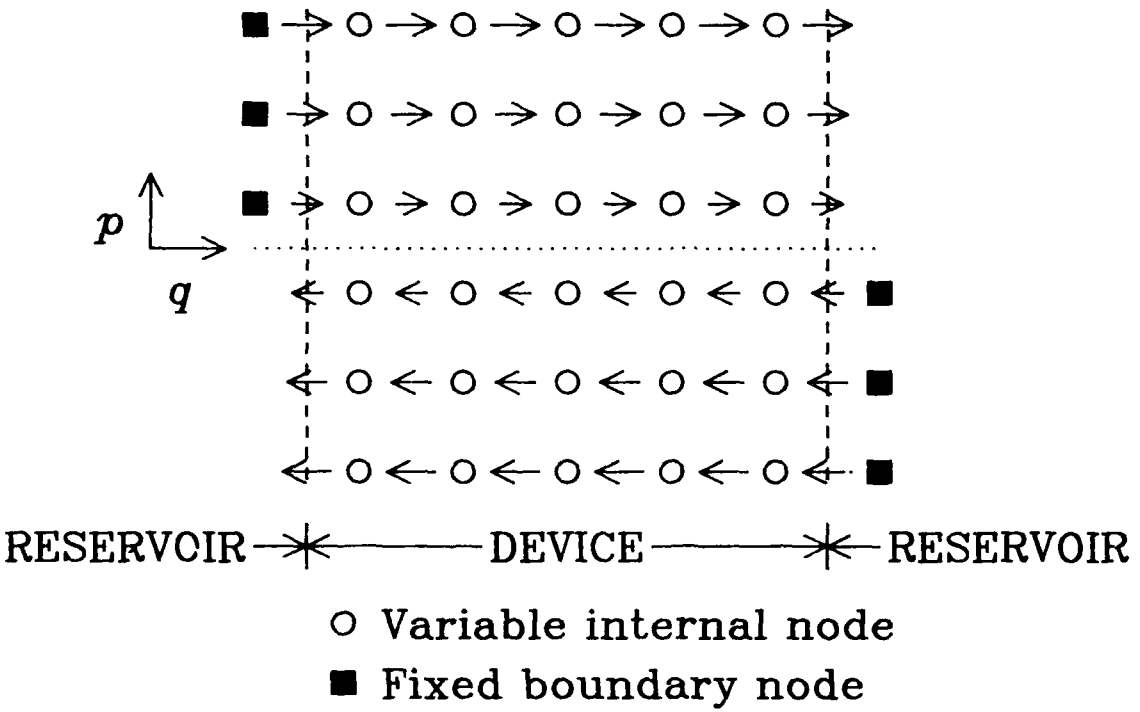

p
q
RESERVOIR
DEVICE
RESERVOIR
Variable internal node
Fixed boundary node

Figure 5

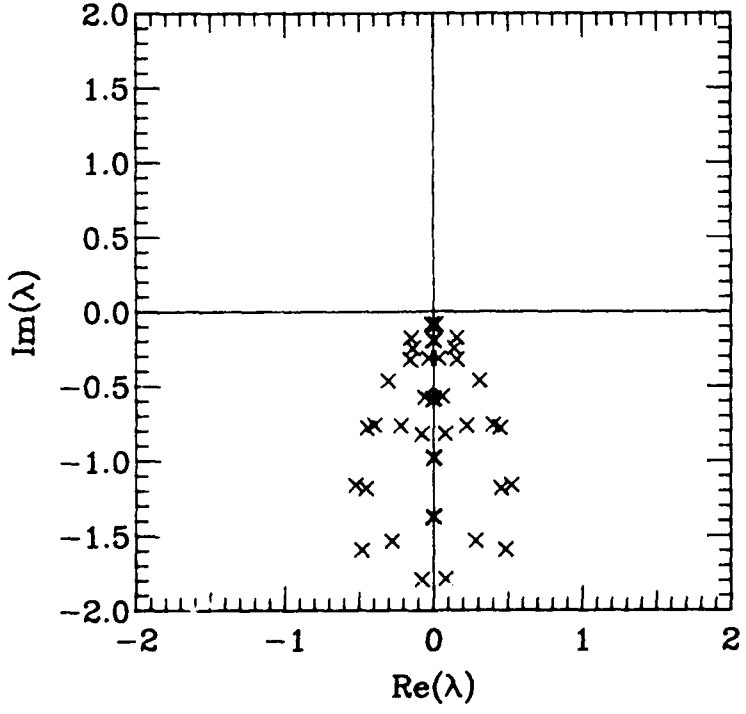


Figure 6

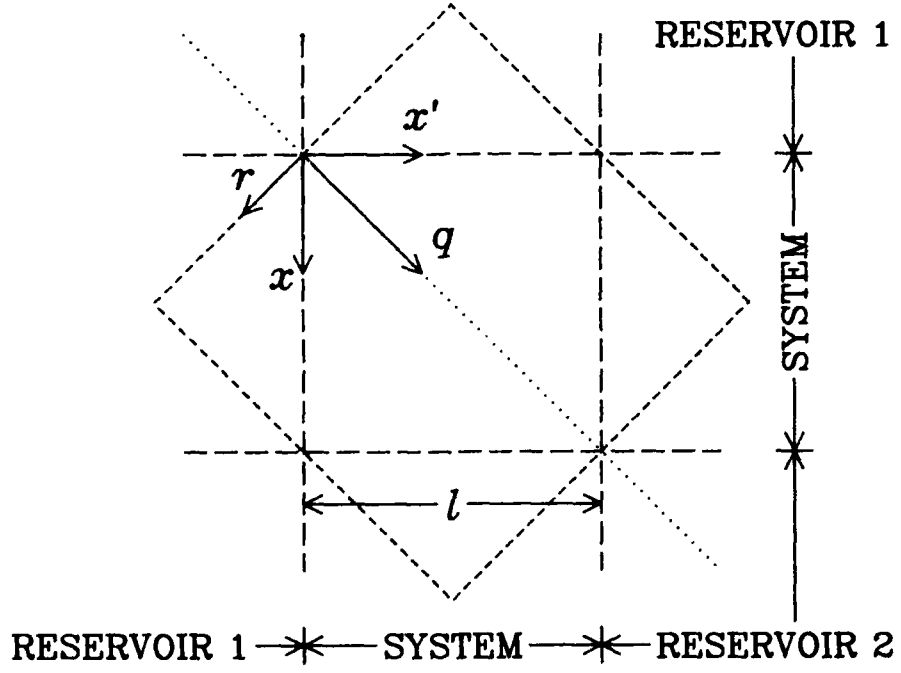


Figure 7

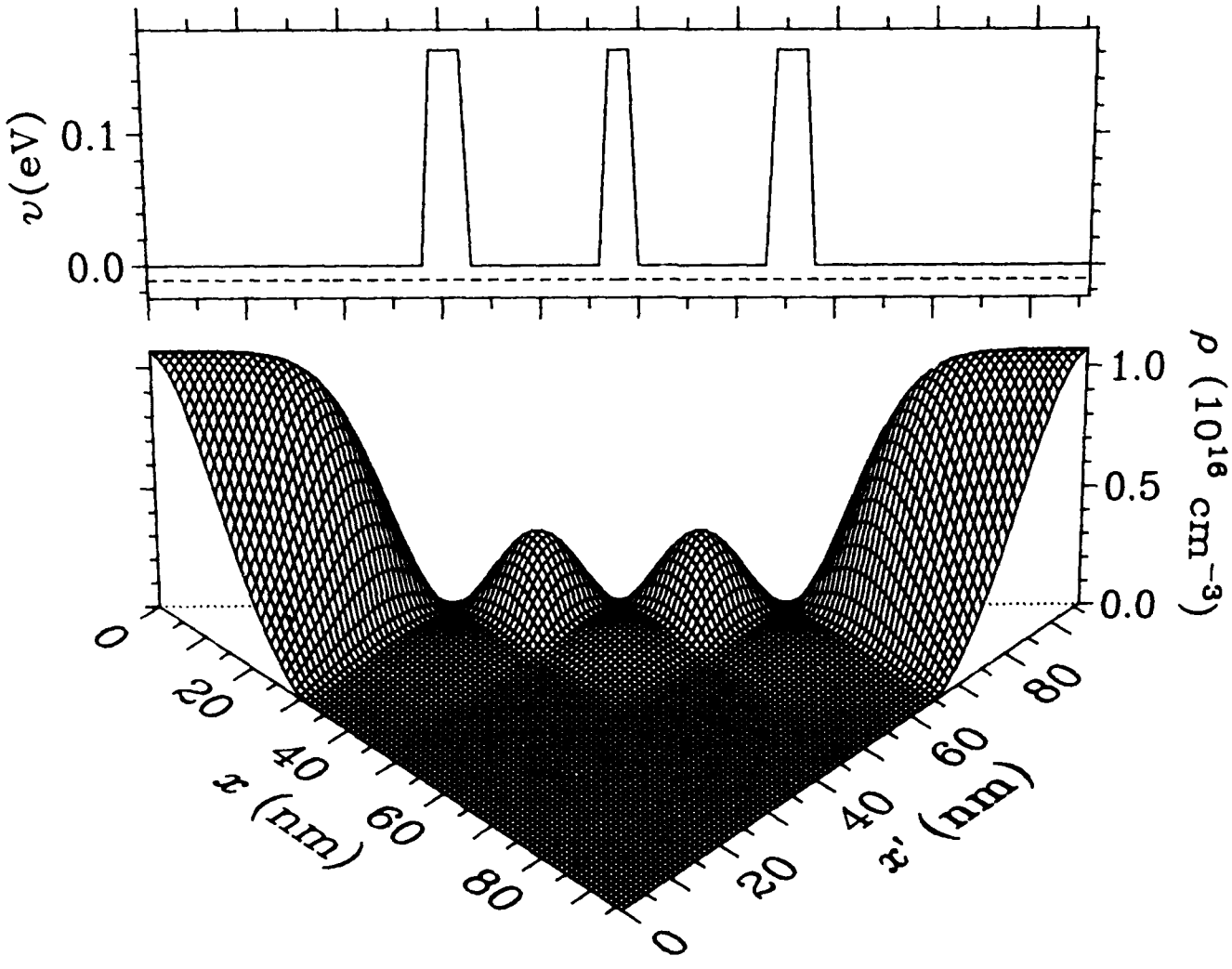


Figure 8

Appendix E

"Vertical Electronic Transport in Novel
Semiconductor Heterojunction Structures"

[Presented at the 1987 International Conference on Superlattices,
Microstructures, and Microdevices, Chicago, August 1987]

VERTICAL ELECTRONIC TRANSPORT IN NOVEL SEMICONDUCTOR HETEROJUNCTION STRUCTURES

Mark A. Reed
Central Research Laboratories
Texas Instruments Incorporated
Dallas, TX 75265 USA

(Received 17 August 1987)

The investigation of vertical transport in semiconductor heterojunction systems has recently undergone a renaissance due to improved epitaxial techniques in a number of material systems. By using resonant tunneling, we can perform electronic spectroscopy not only of the double barrier structure itself, but of any system (with quantized well states) suitably coupled to a resonant tunneling spectrometer. In designing such systems, an important degree of freedom is introduced by utilizing multi-component structures; for example, a GaAs contact - AlGaAs barrier - InGaAs quantum well. In this structure, the high electron affinity of the quantum well creates a "deep" quantum well, in which we demonstrate that quantum well states can be hidden from transport. Finally, we present results from microfabricated quantum well structures ("quantum dots") which are sufficiently small in the lateral dimension to introduce size effects. Telegraph noise due to the lateral size of these structures has been observed, and the first indications of lateral quantization in all three dimensions in a semiconductor quantum well are presented.

1. Introduction

Resonant tunneling in double barrier heterostructures, first investigated in the GaAs/AlGaAs system by Chang, Esaki, and Tsu,[1] has recently been the subject of intense investigation. The remarkable submillimeter wave experiments of Sollner *et al*[2] has generated remarkable interest and success due to improvements in GaAs epitaxial growth techniques. Peak-to-valley tunnel current ratios as large as 3.9:1 at 300K, and 21.7:1 at 77K have been demonstrated in the GaAs/AlAs system.[3] Resonant tunneling by holes,[4] sequential resonant tunneling through a multiquantum well superlattice,[5] double superlattice barrier resonant tunneling structures,[6] and resonant tunneling in triple barrier structures[7] are only among a few of the intriguing investigations that have been performed.

In this paper, we present investigations on vertical transport in a number of new systems primarily to demonstrate the spectroscopic ability of resonant tunneling. The ability to "engineer" the structures allows for interesting electronic spectroscopy, previously available only with optical techniques. We also present results on transport in submicron structures that are sufficiently small in the lateral dimensions (as well as in the vertical dimension) such that size effects are observable.

2. Review: Quantum Well Spectroscopy and Deep Quantum Wells

The phenomena of resonant tunneling is characterized by the appearance of negative differential resistance regions in the I-V characteristics due to resonant tunneling through the quantum well state of the double barrier structure. Figure 1 illustrates a typical example of resonant tunneling in a GaAs/AlGaAs structure grown by MBE. Experimental details have been described previously.[6] In this case the quantum well width and tunnel barrier height have been adjusted such that there are two states in the well (ground and first excited state) that will be available to transport with the structure under bias. The two well-defined peaks in the characteristics of Figure 1 correspond to these resonances. The calculated peak positions of the structure (83 meV for the ground state and 350 meV for the first excited state) are in excellent

agreement with the experimentally observed peaks, when a series resistance of 40Ω is taken into account. Although suitably designed structures (like the one illustrated here) are easily fit by an effective series resistance. a more precise determination of the peak positions demand a detailed modeling of the accumulation and depletion regions in the contacts. This is especially important in view of the recent trend in these structures; i.e., inserting undoped GaAs "spacers" between the doped GaAs contact and the AlGaAs barrier to prevent Si diffusion into the double barrier region.. Figure 2 shows the shifting of the resonant peak to higher voltages with increasing GaAs spacer thickness, as expected.

The design of these structures need not be limited to the simple "double square barrier" case; in addition to being able to substitute complex barrier shapes,[6] it has been demonstrated[8] that the central quantum well can be replaced by a material with higher electron affinity than the surrounding GaAs contact regions, specifically, a strained-layer InGaAs quantum well. Figure 3 illustrates the gradual shifting of the ground state resonance to lower voltages as In is incorporated into the well of a nominally identical resonant tunneling structure. It should be noted that the lowest voltage structure (c) has a finite zero-bias conductance since the quantum well state has been lowered below the Fermi level of the GaAs contacts.

Figure 4 illustrates the effect of adding sufficient In to the quantum well that not only lowers the quantum well ground state below the Fermi level, but below the GaAs contact conduction band edge. In this case, the ground state is hidden from transport and resonant tunneling proceeds only through the excited states. Notice here that the conduction band edge of the InGaAs was lowered sufficiently to observe the $n=3$ state of the quantum well, which was virtual in the GaAs quantum well case.

3. Tunneling from Quantized Regions

The ability to engineer peak positions in multicomponent systems now allows us, in general, to use double barrier structures as spectrometers on suitably designed quantum well or superlattice structures. Specifically, we demonstrate here that we can perform spectroscopy on large quantum wells placed between different GaAs (or InGaAs) quantum well, double barrier structures. The different resonant

positions of the peaks, tunable by In content instead of changing the quantum well width (which introduces other complexities), discriminates which region of the epitaxial structure is being examined. This insertion of these resonant tunneling regions throughout a complex epitaxial structure gives us a microscopic spectrometer to examine the energy states in the structure.

Figure 5 illustrates an experimental embodiment of such a structure. The structure consists of a 40Å $Al_{.25}Ga_{.75}As$ double barrier (90Å $In_{.05}Ga_{.95}As$ quantum well) / 750Å GaAs quantum well / 40Å $Al_{.25}Ga_{.75}As$ double barrier (90Å GaAs quantum well), with a n^+ GaAs contact on the other side of the GaAs quantum well resonant tunneling structure. The large GaAs quantum well (750Å) will have a small energy splitting in comparison to the quantum wells of the double barrier structures (90Å). In this specific case, the GaAs quantum well resonant tunneling structure has a n^+ contact region so as to provide a "source" or "sink" of available carriers. Thus, when the structure is biased such that carriers tunnel from the large quantum well into the double barrier structure, they will be injected from a ladder of states in the large quantum well. For simplicity, we will discuss results using the GaAs quantum well resonant tunneling structure as the spectrometer (though similar results are also obtained from the InGaAs resonant tunneling structure, at a different bias position). In addition, the heavily doped contact on the right hand side and the intrinsic region on the left hand side (in Fig. 5(b)) create accurately the band structure schematically illustrated in Figure 5; a more detailed self-consistent solution of the band structure is remarkably similar. Thus, the intrinsic region simply acts as a lever arm for the applied voltage, as well as being quantized.

Figure 6(a) shows a I-V characteristic at $T = 4.2K$ corresponding to electron injection from the large GaAs quantum well. A series of peaks corresponding to electron injection from the states in the 750Å quantum well through the state in the 90Å quantum well are clearly observable. The structure appears on the low bias side of the major peak only. The experimentally observed splittings are 59 meV, 103 meV, 143 meV, and 238 meV. No peaks corresponding to $n > 4$ in the large quantum well are observed, indicative of the position of the Fermi level in this region. The ratios of the splittings to the ground state splitting (1:1.74:2.42:4.03) are in excellent agreement with

calculated values (1:1.69:2.36:3.01) except for the n=4 level, presumably due to band bending. Clearly this technique can be generalized to more complex regions, such as parabolic injectors to verify equal splitting in such structures.

Now consider electron injection from the n+ GaAs region, through the 90Å GaAs quantum well double barrier structure, into the 750Å GaAs quantum well region. Below the resonant peak, the tunneling current is a sum of elastic scattering to available states on the other side of the structure, and inelastic tunneling through the entire structure or through the intermediate quantum well state (which should be negligible at these temperatures). Thus, no structure in the I-V characteristic below the resonant peak should be seen due to the large quantum well. However, when the structure is biased beyond the major resonant peak position, elastic scattering can then occur via the 90Å quantum well state; i.e., elastic scattering to the 90Å quantum well ground state subband, relaxation to the bottom of this band, then tunneling out of the structure. Thus, peaks in the I-V characteristic should occur at biases greater than the major resonant peak bias.

Figure 6(b) shows the I-V characteristic of this structure at T=4.2K corresponding to electron injection from the n+ GaAs region. There are no observable peaks at biases less than the resonant bias. However, a series of oscillations appear at biases greater than the resonant bias in accordance with the mechanism described above. The spacings of these oscillations (~130meV) are approximately constant, corresponding to the asymptotic energy level spacing of a large quantum well, and quantitatively are in good agreement with the size of the splittings of these levels when the asymmetry of the resonant peak positions is taken into account. However, there is insufficient resolution of the splitting to allow an exact determination of the quantum numbers of these ' vels.

4. Transport through Quantum Dots

The resonant tunneling devices discussed so far operate by virtue of the quantum size effect in the quantum well imposed by the lower electron affinity tunnel barriers. The dimension of the structure in the plane of the quantum well is essentially infinite on this scale. However, microfabrication techniques have advanced to the degree that lateral dimensions can approach the dimensions determined by the epitaxial growth

layers. In this case, splitting of the quantum well state (band) will occur due to lateral quantization imposed by a microfabricated potential. We have produced such structures, and have measured transport through these laterally confined quantum wells, i.e. "quantum dots".

The microfabrication approach used to produce these microstructures is summarized in Figure 7. The initial structure (substrate) is a resonant tunneling diode structure, grown by MBE on a n^+ GaAs substrate and consisted of a 0.5 micrometer-thick Si-doped ($2x10^{18}$ cm^{-3}) GaAs buffer layer graded to less than 10^{16} cm^{-3}, a 50Å undoped GaAs spacer layer, a 50Å $Al_{0.27}Ga_{0.73}As$ tunnel barrier, and a 50Å undoped GaAs quantum well. The structure was grown to be nominally symmetric about a plane through the center of the quantum well. Large area (2 micrometers x 2 micrometers) devices fabricated in a conventional manner exhibit a 1.6:1 peak to valley tunnel current ratio and a current density at resonance of 1.6×10^4 A/cm^{-2}.

E-beam lithography was used to define an ensemble of quantum dots (including single and multiple dot regions) nominally 0.25, 0.15, and 0.1 micrometer in diameter, in a bi-layer PMMA resist spun onto the structure. This pattern was then transferred to a AuGe/Ni/Au (500Å / 150Å / 600Å) dual-purpose Ohmic top contact and etch mask by lift-off. Highly anisotropic reactive ion etching (RIE) using BCl_3 as an etch gas defined columns in the structure. A SEM of one of these etched structures is seen in Figure 8. To make contact to the tops of the column(s), an insulating polyimide was spun on the wafer, cured, and then etched back by oxygen RIE until the tops of the columns were exposed. A gold contact layer was lifted off which provided bonding pads for contact to either single dots or arrays of them.

Figure 9 shows a I-V characteristic of a single quantum dot resonant tunneling structure at 100K. The lateral dimensions of this single dot structure is 0.15 micrometers x 0.25 micrometers. The structure clearly shows NDR, though the peak to valley is degraded from the large area structure, probably due to process damage. The I-V clearly exhibits a "noise" that is far above the system background noise. The origin of this noise is the so-called "single electron switching" phenomena[9] that has been observed in narrow Si MOSFET wires. Traps in or near the narrow conduction channel and are near the Fermi level can emit or capture electrons with a temperature-dependent characteristic time. The lowering of specific traps

through the Fermi level is clearly evident (at .6V, .8V through .85V, 1.0V through 1.1V). This phenomena is seen up to room temperature (though for a different set of traps), and is usually "frozen out" by 4.2K.

Figure 10 shows a time-dependent trace of a similar device at fixed bias voltage. The switching between two discrete states is evident. In all regions of Figure 9 where a trap is biased near the Fermi level, switching similar to Figure 10 is observable, provided that the temperature is in the correct range. The switching rates for different traps are in general unique, and are dependent on temperature and bias. Under the appropriate temperature/bias conditions, a superposition of these phenomena can be seen (for example, see Fig. 9, ~ 1.1V).

If the mechanism for the telegraph noise is scattering from traps of varying occupancy as suggested, then the trap(s) should have a well-defined activation energy measurable by the switching rate between discrete values as a function of temperature. This was measured by taking the same device as shown in under constant bias (specifically, the same as for Figure 10) and varying the temperature. Figure 11 shows that the behavior is indeed activated, with a measured activation energy of of 280 meV for this particular trap.

It is curious that telegraph noise can be seen for the physical dimensions (0.15 micrometer x 0.25 micrometer) of the structure.[9] However, the effects of depletion at the etched mesa side surfaces due to pinning of the Fermi level has not been taken into account. Taking the observed current at resonance and assuming that the current density must be the same as in the large area device (assuming that the switching is a perturbation), we calculate that the effective (circular) conduction path diameter is ~500Å, consistent with the observation of the switching phenomena. This implies a depletion layer of approximately 500Å. However, transport and switching phenomena have also been observed in an array of dots 2000Å by 1000Å suggesting a depletion layer smaller than 500Å.

At these lateral dimensions, splitting of the quantum well resonance due to lateral size quantization imposed by the depletion depth potential should be observable. Very recent results of one of these structures, epitaxially similar with lateral dimensions of 1000Å round, is shown in Figure 12. The I-V characteristics show well defined negative differential resistance peaks at low temperature, which disappear at high

temperature. This is due to resonant tunneling through the 0-D density of states in the quantum "dot", and is the first indication of lateral quantization in all three dimensions in a semiconductor quantum well. Detailed measurements of this structure are in progress.[10]

5. Summary

The range of new phenomena in the vertical transport of heterostructure and microfabricated heterostructure systems is rapidly expanding. The above provides some techniques useful for the design and spectroscopy that is available in these structures. We also have described the techniques for the fabrication of microstructured multilayers, which already show interesting size effect phenomena and the first indication of lateral size quantization to all three dimensions in a semiconductor quantum well. These structures should provide a unique laboratory for the investigation of localization and quantum size effect phenomena.

Acknowledgments - I am indebted to J. W. Lee and R. J. Matyi for MBE growth, to J. N. Randall for quantum dot microfabrication, to T. M. Moore for e-beam lithography, to H.-L. Tsai for cross-section TEMs, to R. T. Bate, W. R. Frensley and A. E. Wetsel for analysis and discussions, and to R. K. Aldert, D. A. Schultz, P. F. Stickney, J.R. Thomason, A. E. Wetsel, and J. A. Williams for technical assistance. This work was sponsored by the Office of Naval Research, the Army Research Office, and the Air Force Wright Aeronautical Laboratory.

References

1. L. L. Chang, L. Esaki, and R. Tsu, *Applied Physics Letters* **24**, 593 (1974).
2. T.C.L.G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, *Applied Physics Letters* **43**, 588 (1983).
3. C. I. Huang, M. J. Paulus, C. A. Bozada, S. C. Dudley, K. R. Evans, C. E. Stutz, R. L. Jones, and M. E. Cheney, *Applied Physics Letters* **51**, 121 (1987).
4. E. E. Mendez, W. I. Wang, B. Ricco, and L. Esaki, *Applied Physics Letters* **47**, 415 (1985).
5. F. Capasso, K. Mohammed, and A. Y. Cho, *Applied Physics Letters* **48**, 478 (1986).
6. M. A. Reed, J. W. Lee, and H-L. Tsai, *Applied Physics Letters* **49**, 158 (1986).
7. T. Nakagawa, T. Fujita, Y. Matsumoto, T. Kojima, and K. Ohta, *Applied Physics Letters* **51**, 445 (1987).
8. M. A. Reed and J. W. Lee, *Applied Physics Letters* **50**, 845 (1987).
9. K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, *Physics Review Letters* **52**, 228 (1984).
10. M. A. Reed, J. N. Randall, R. J. Aggarwal, R. J. Matyi, T. M. Moore, and A. E. Wetsel, *Phys. Rev. Lett.*, to be published.

FIGURE CAPTIONS

Figure 1. I-V characteristics of a 100Å GaAs quantum well / double $Al_{.3}Ga_{.7}As$ barrier structure at 77K. (a) Characteristic demonstrating the ground and first excited state. (b) An expanded scale showing the ground state resonance.

Figure 2. Resonant voltage position (T=77K) as a function of GaAs undoped spacer layer thickness for a 50Å GaAs quantum well / 50Å $Al_{.3}Ga_{.7}As$ double barrier structure. The spacer lies between the AlGaAs barrier and the n^+ GaAs contact ($1x10^{18}$ cm^{-3}), which was graded to $1x10^{16}$ cm^{-3} over ~200Å prior to the spacer.

Figure 3. I-V characteristics of a 50Å $In_yGa_{1-y}As$ quantum well / double 40Å $Al_{.25}Ga_{.75}As$ barrier structure at 77K. (a) y=0 (right hand current scale). (b) y=0.03 (left hand current scale). (c) y=0.08 (left hand current scale).

Figure 4. (a) I-V characteristics of a 85Å GaAs quantum well / double 35Å AlAs barrier structure at 77K. The ground and first excited state are visible. The inset is an expanded view of the ground state resonance. (b) I-V characteristics of a 85Å $In_{.1}Ga_{.9}As$ quantum well / double 35Å AlAs barrier structure at 77K. The ground state is hidden, and the first (n=2) and second (n=3) excited states are visible. Note the absence of the low bias resonance seen in (a).

Figure 5. (a) Schematic of the large GaAs quantum well / resonant tunneling spectrometer structure. (b) Bias condition for injection from the large GaAs quantum well.

Figure 6. (a) I-V characteristic at T=4.2K of the large GaAs quantum well / double barrier structure, corresponding to electron injection from the 750Å GaAs quantum well. The structure on the low bias side of the major peak is due to tunneling from the levels in the 750Å GaAs quantum well. (b) I-V characteristic at T=4.2K of the large GaAs quantum well / double barrier

structure, corresponding to electron injection from the n+ GaAs contact. The structure on the high bias side of the major peak is due to tunneling into the levels in the 750Å GaAs quantum well.

Figure 7. Schematic of fabrication sequence of quantum dot devices. (a) GaAs/AlGaAs double barrier structure starting material, with n+ GaAs contacts on top and bottom. (b) E-beam definition in PMMA of dots (singular or multiple; 3 shown here). Evaporation of AuGe/Ni/Au Ohmic contacts. (c) Liftoff. BCl_3 reactive ion etch to define pillars with "dots". (d) Planerization with polyimide. (e) Etchback to Ohmic top contacts. Evaporation of Au bonding pads.

Figure 8. Scanning electron micrograph of a single anisotropically etched column containing a quantum dot. The marker is 0.5 micrometer.

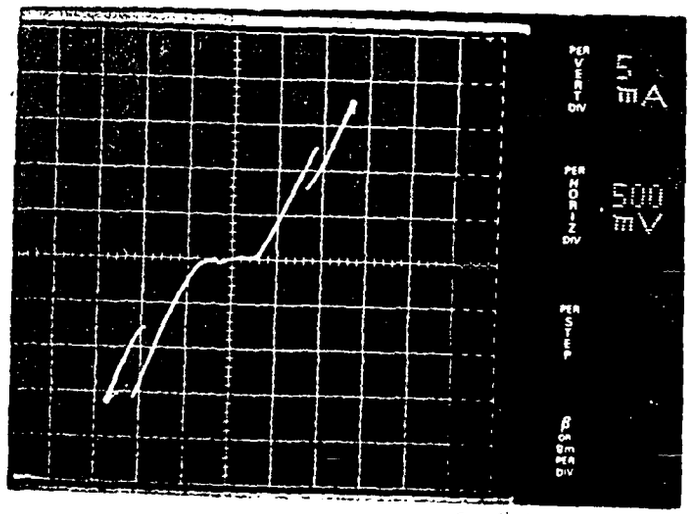
Figure 9. I-V characteristics at 100K of a single quantum dot which has lateral dimensions 0.25 x 0.15 micrometer.

Figure 10. Time dependent resistance fluctuations at a fixed bias of a single quantum dot. T = 180K.

Figure 11. Measurement of the activation energy of the trap responsible for the telegraph noise seen in Figure 10.

Figure 12. I-V characteristic of a single quantum dot nanostructure at 4.2K, showing resonant tunneling through the discrete states of the quantum dot.

(a)


PER VERT DIV
5 mA
PER HORIZ DIV
500 mV
PER STEP
BEAM PER DIV

(b)

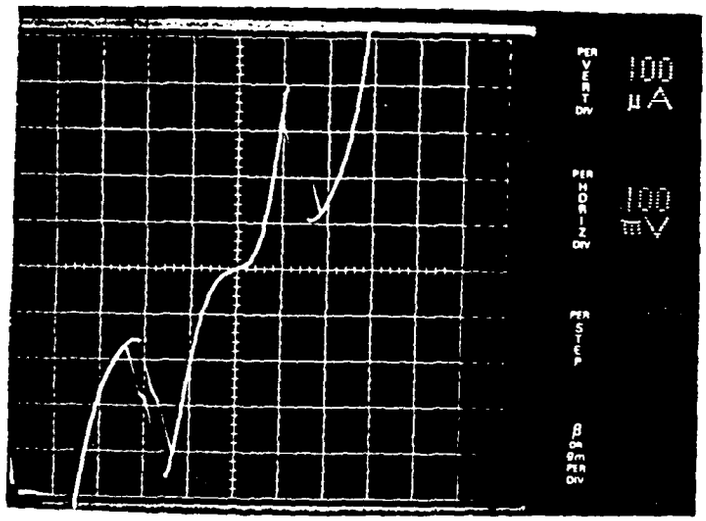

PER VERT DIV
100 µA
PER HORIZ DIV
100 mV
PER STEP
BEAM PER DIV

FIG 1

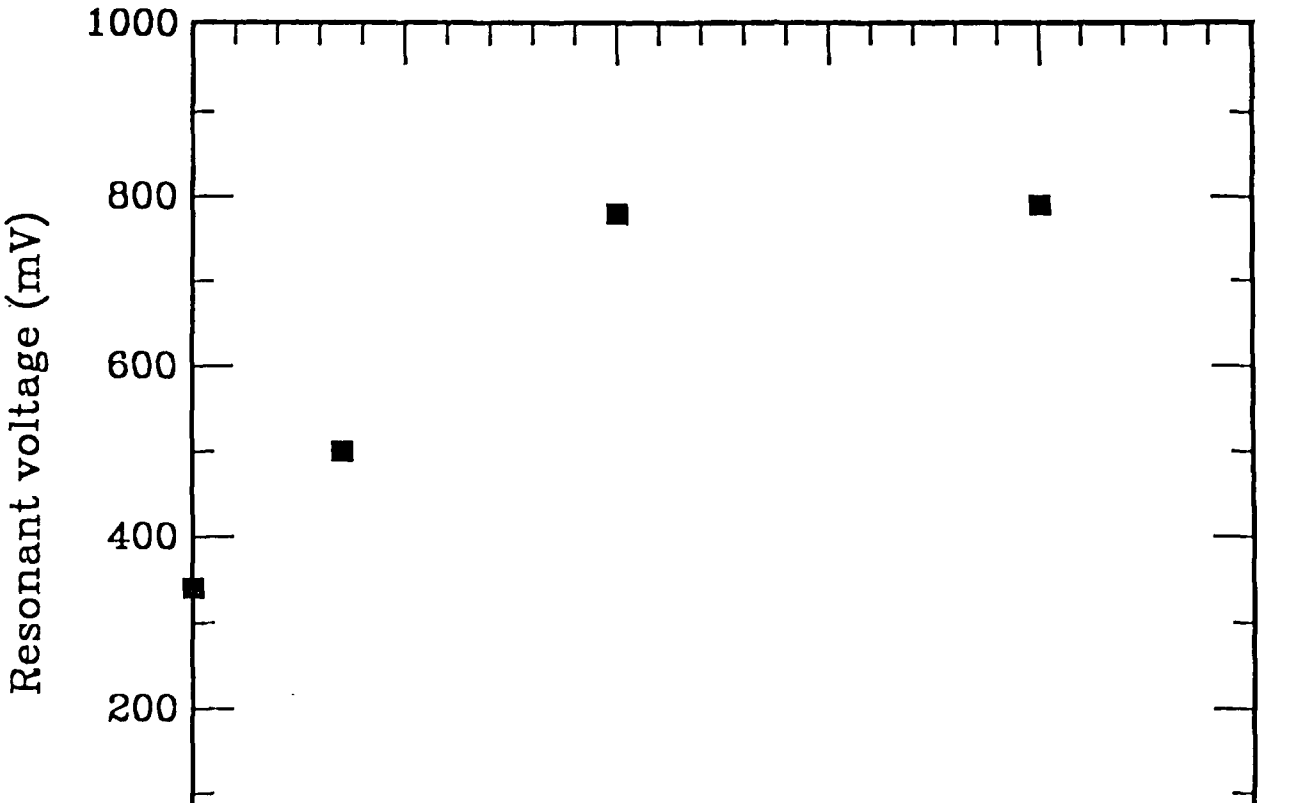

1000
800
600
400
200
0
Resonant voltage (mV)
0
50
100
150
200
250
Spacer layer thickness (A)

FIG 2

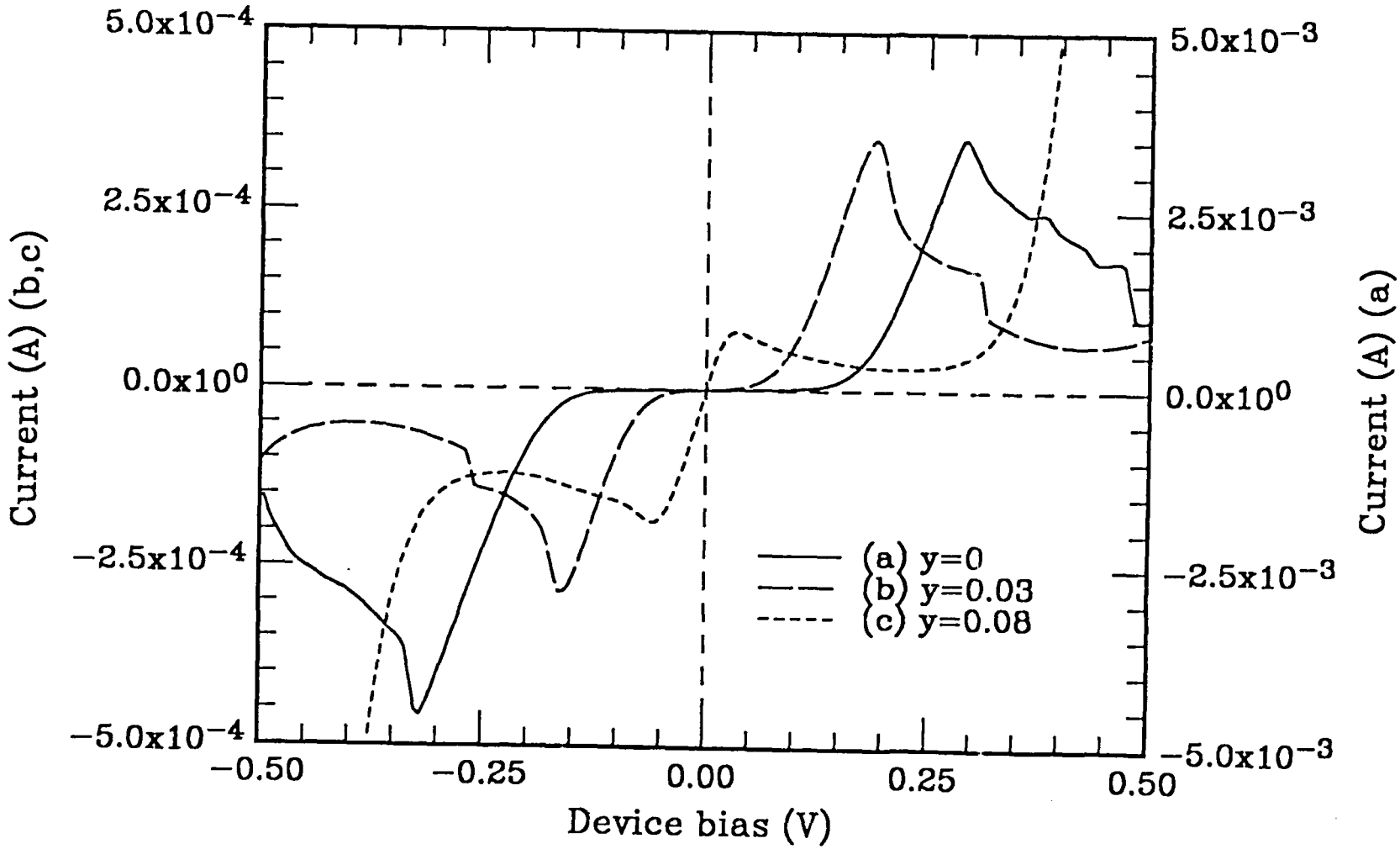
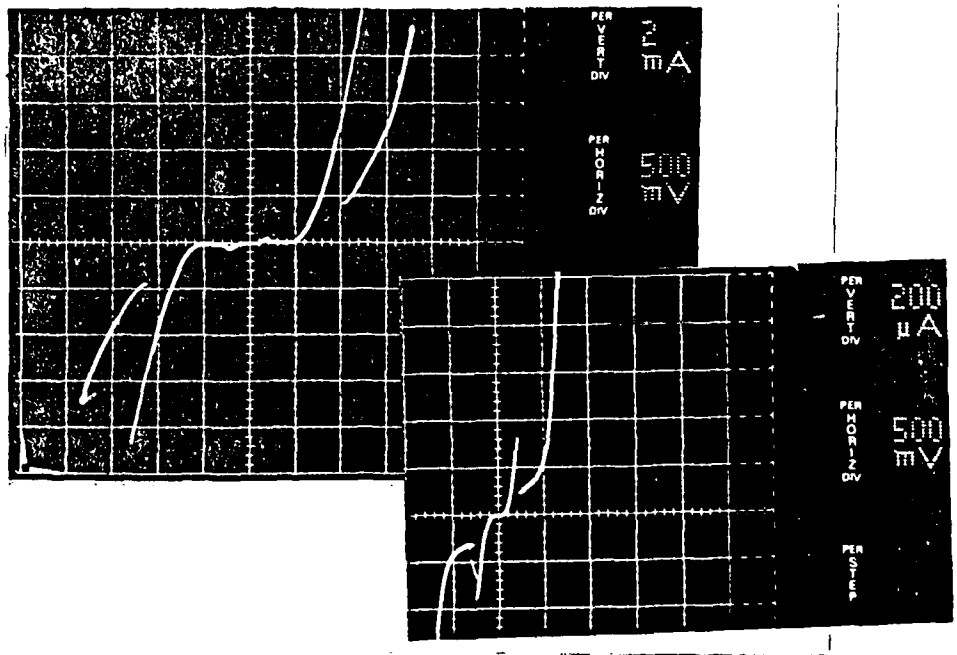

5.0x10⁻⁴
2.5x10⁻⁴
0.0x10⁰
−2.5x10⁻⁴
−5.0x10⁻⁴
Current (A) (b,c)
5.0x10⁻³
2.5x10⁻³
0.0x10⁰
−2.5x10⁻³
−5.0x10⁻³
Current (A) (a)
−0.50
−0.25
0.00
0.25
0.50
Device bias (V)
(a) y=0
(b) y=0.03
(c) y=0.08

FIG. 3

(a)


PER VERT DIV
2 mA
PER HORIZ DIV
500 mV
PER VERT DIV
200 µA
PER HORIZ DIV
500 mV
PER STEP

(b)

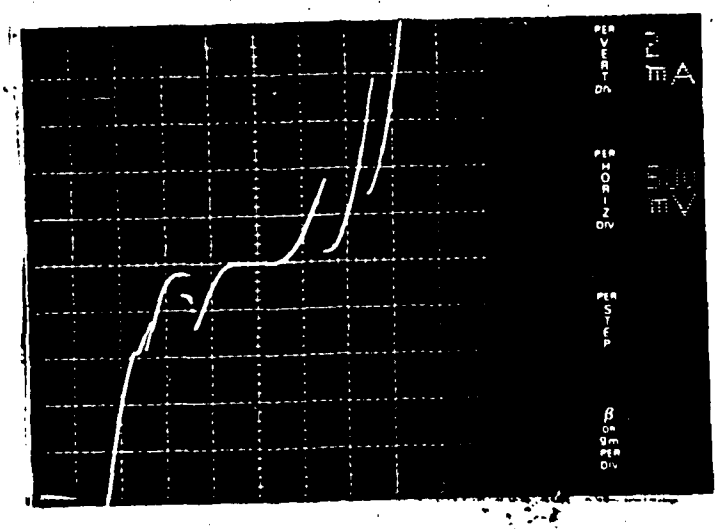
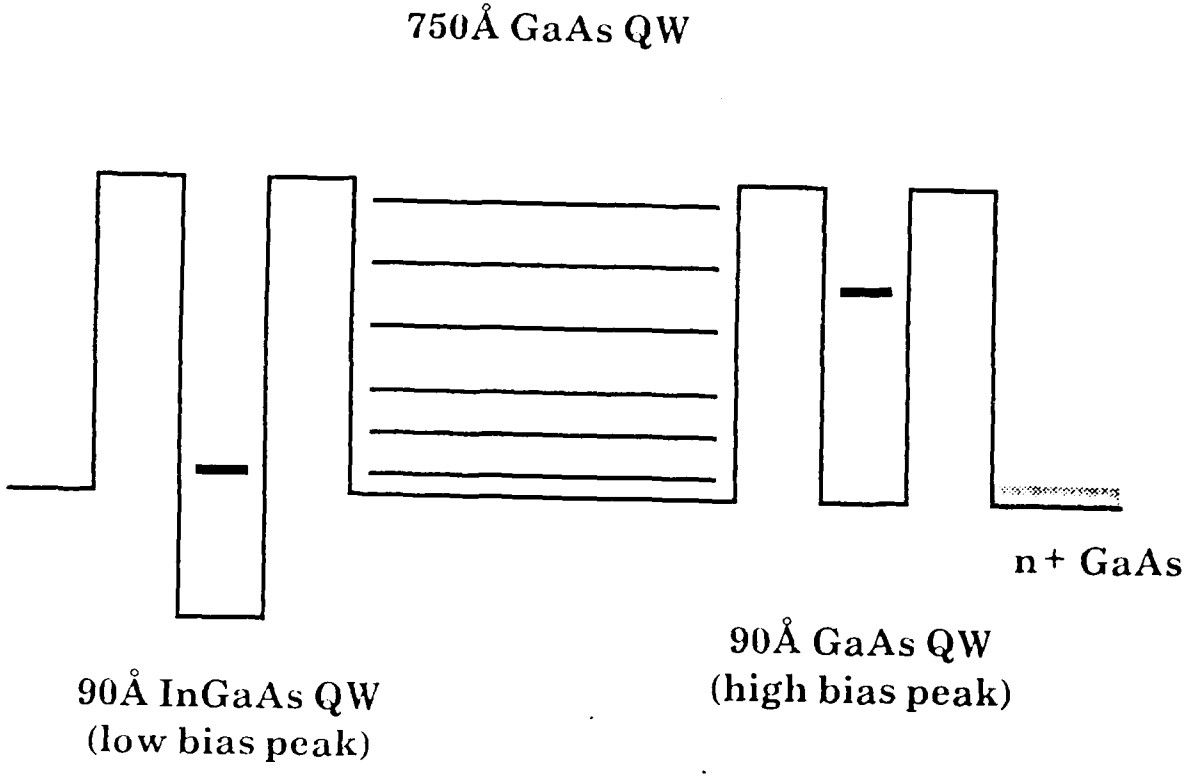
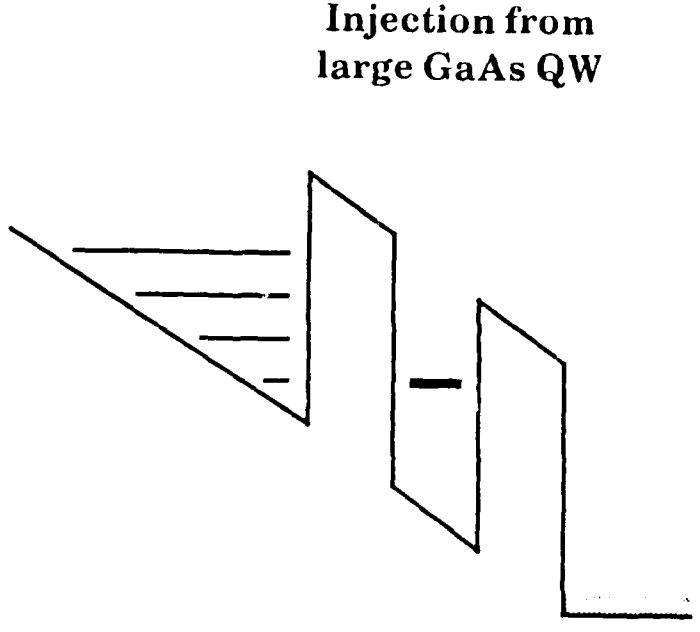

PER VERT DIV
2 mA
PER HORIZ DIV
500 mV
PER STEP

FIG. 4


750Å GaAs QW
n+ GaAs
90Å InGaAs QW
(low bias peak)
90Å GaAs QW
(high bias peak)

(a)


Injection from
large GaAs QW

(b)

FIG. 5

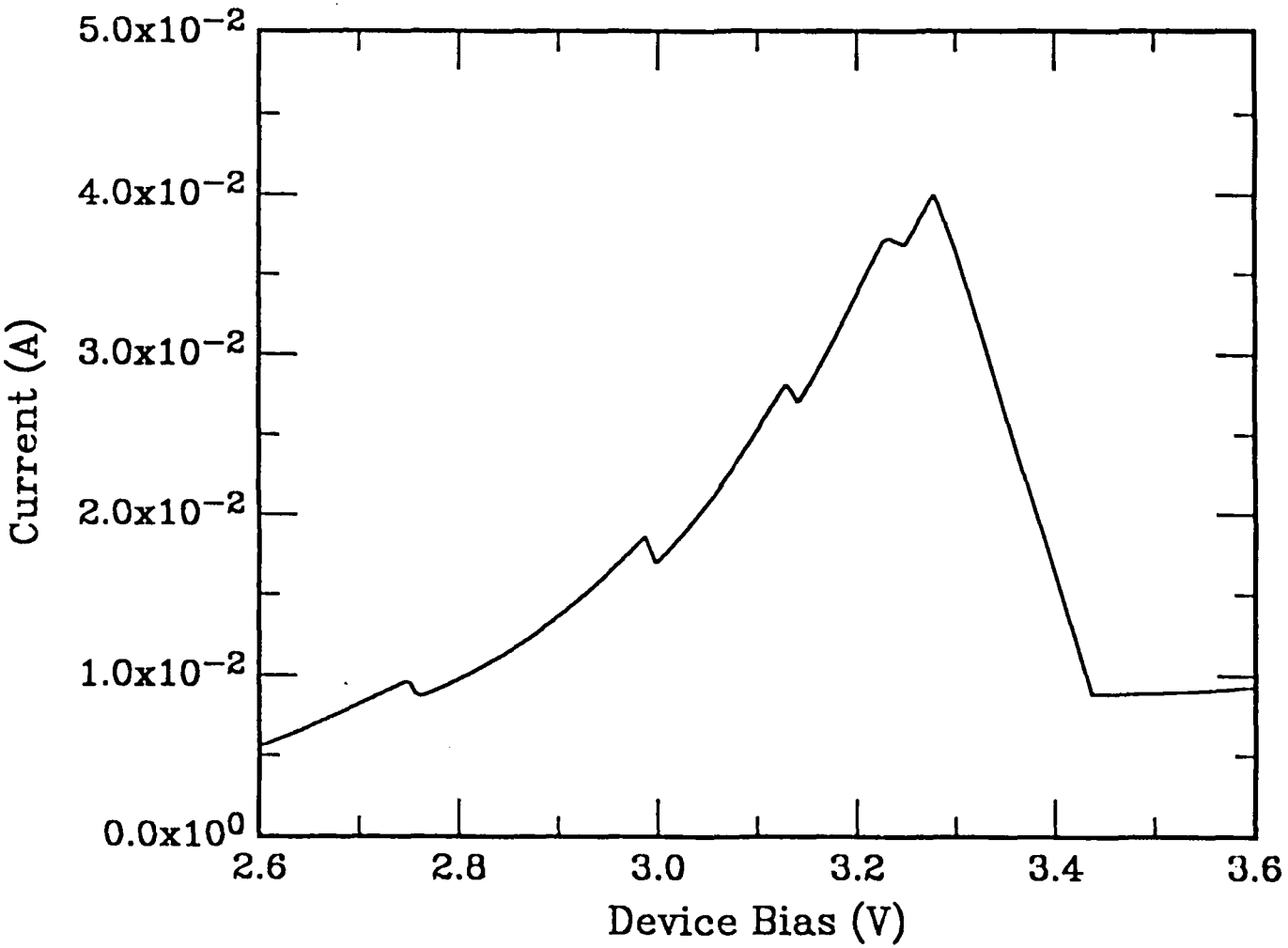

5.0x10^{-2}
4.0x10^{-2}
3.0x10^{-2}
2.0x10^{-2}
1.0x10^{-2}
0.0x10^{0}
Current (A)
2.6
2.8
3.0
3.2
3.4
3.6
Device Bias (V)

FIG. 6(a)

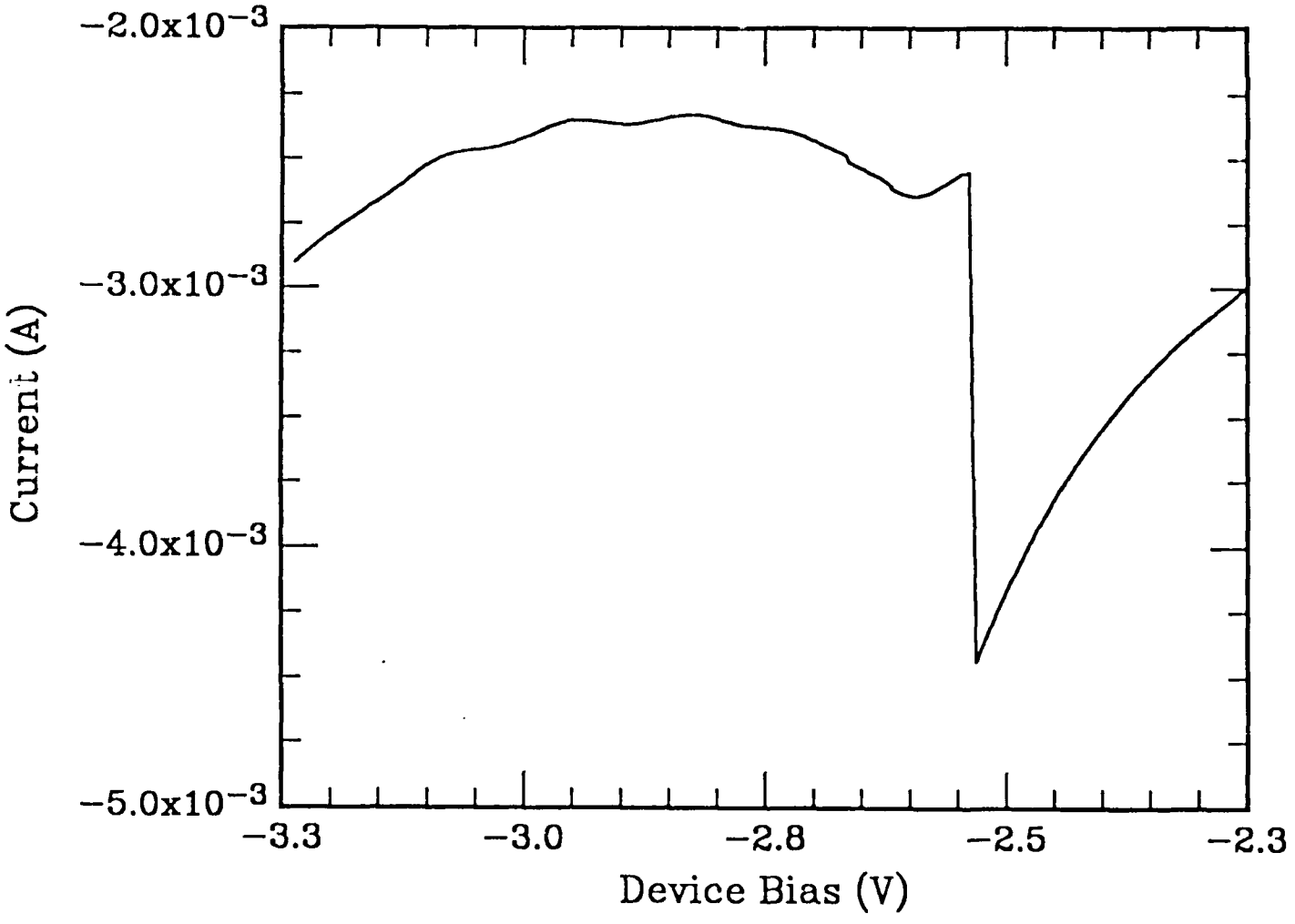
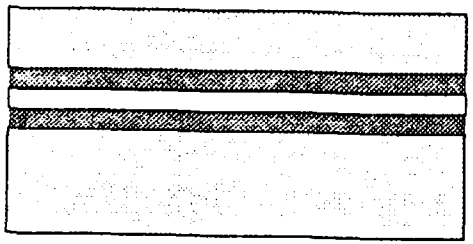
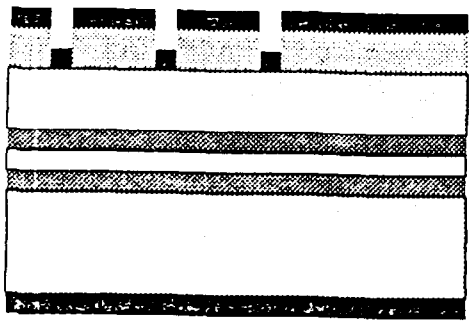

−2.0x10−3
−3.0x10−3
−4.0x10−3
−5.0x10−3
Current (A)
−3.3
−3.0
−2.8
−2.5
−2.3
Device Bias (V)

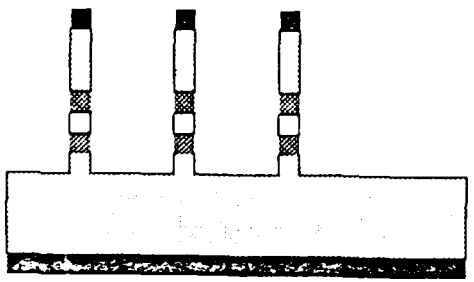
FIG. 6(b)



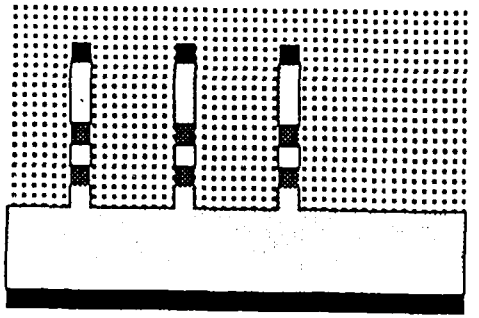
(a)



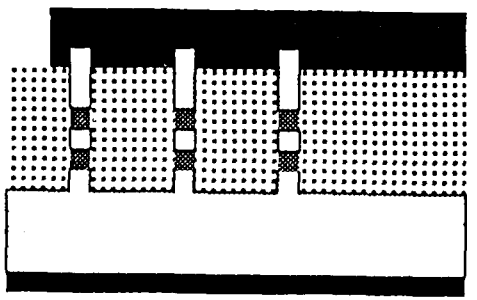
(b)



(c)



(d)



(e)

FIG. 7

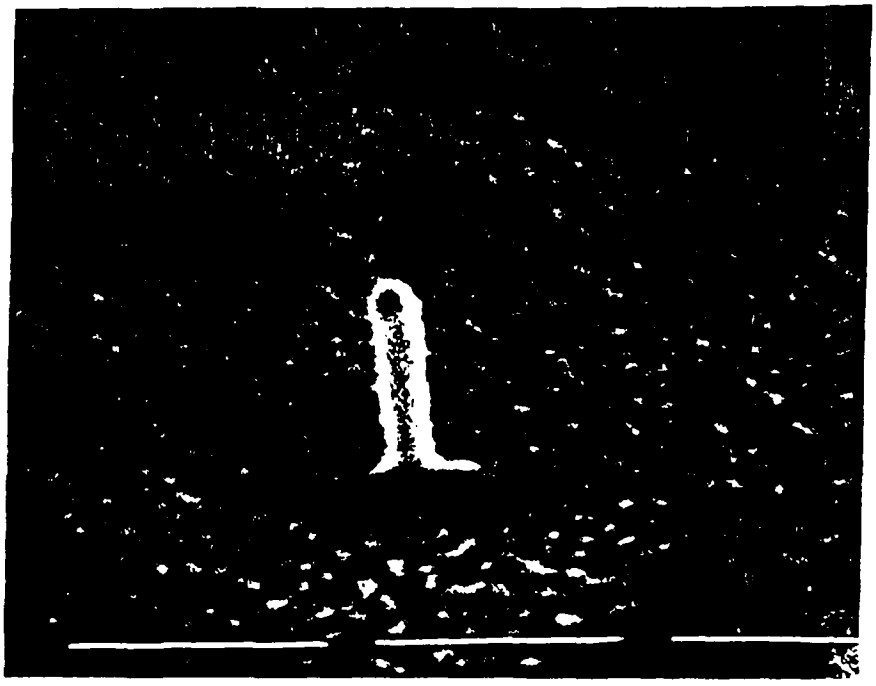


FIG. 8

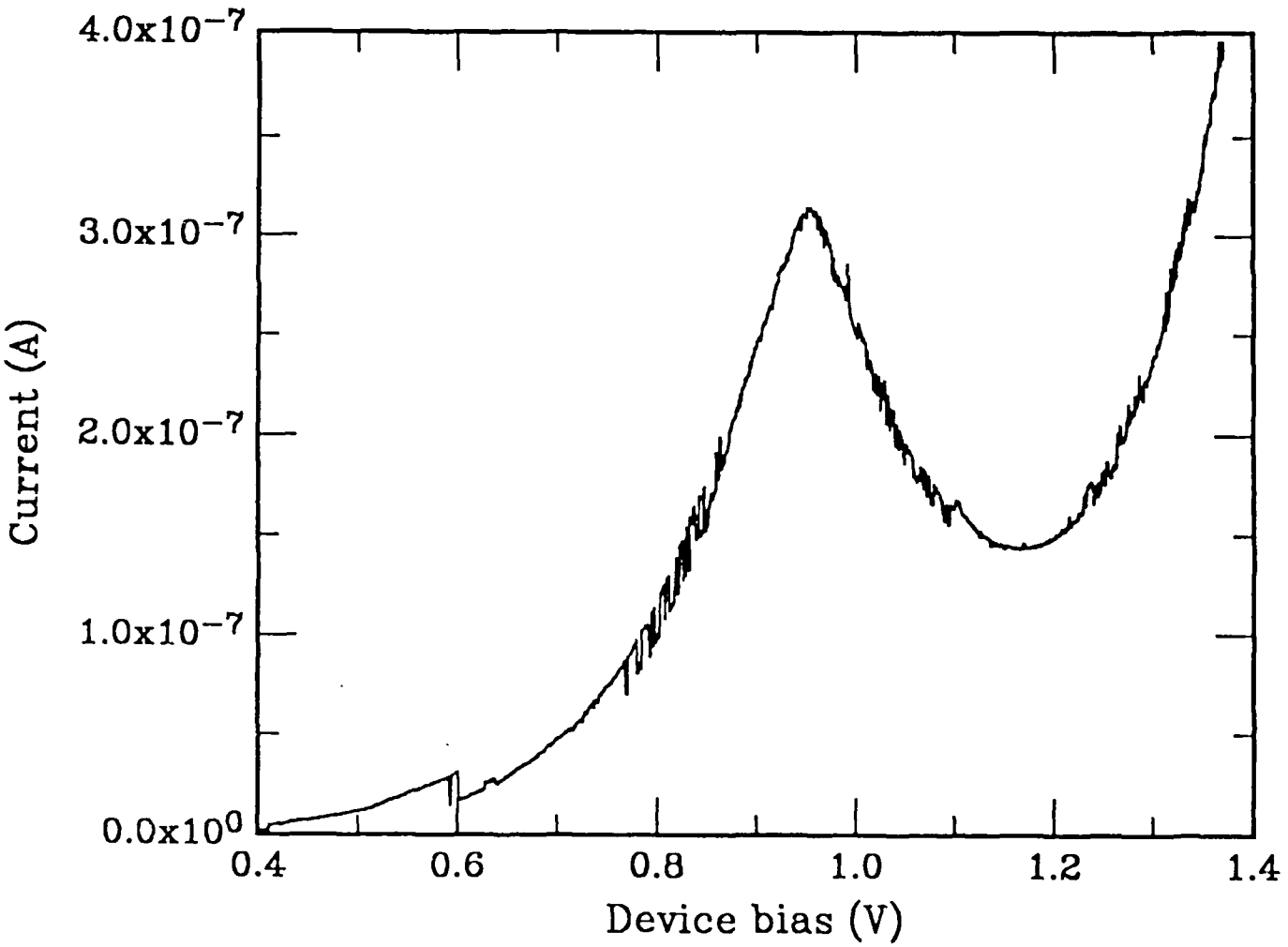

4.0x10^{-7}
3.0x10^{-7}
2.0x10^{-7}
1.0x10^{-7}
0.0x10^{0}
Current (A)
0.4
0.6
0.8
1.0
1.2
1.4
Device bias (V)

FIG. 9

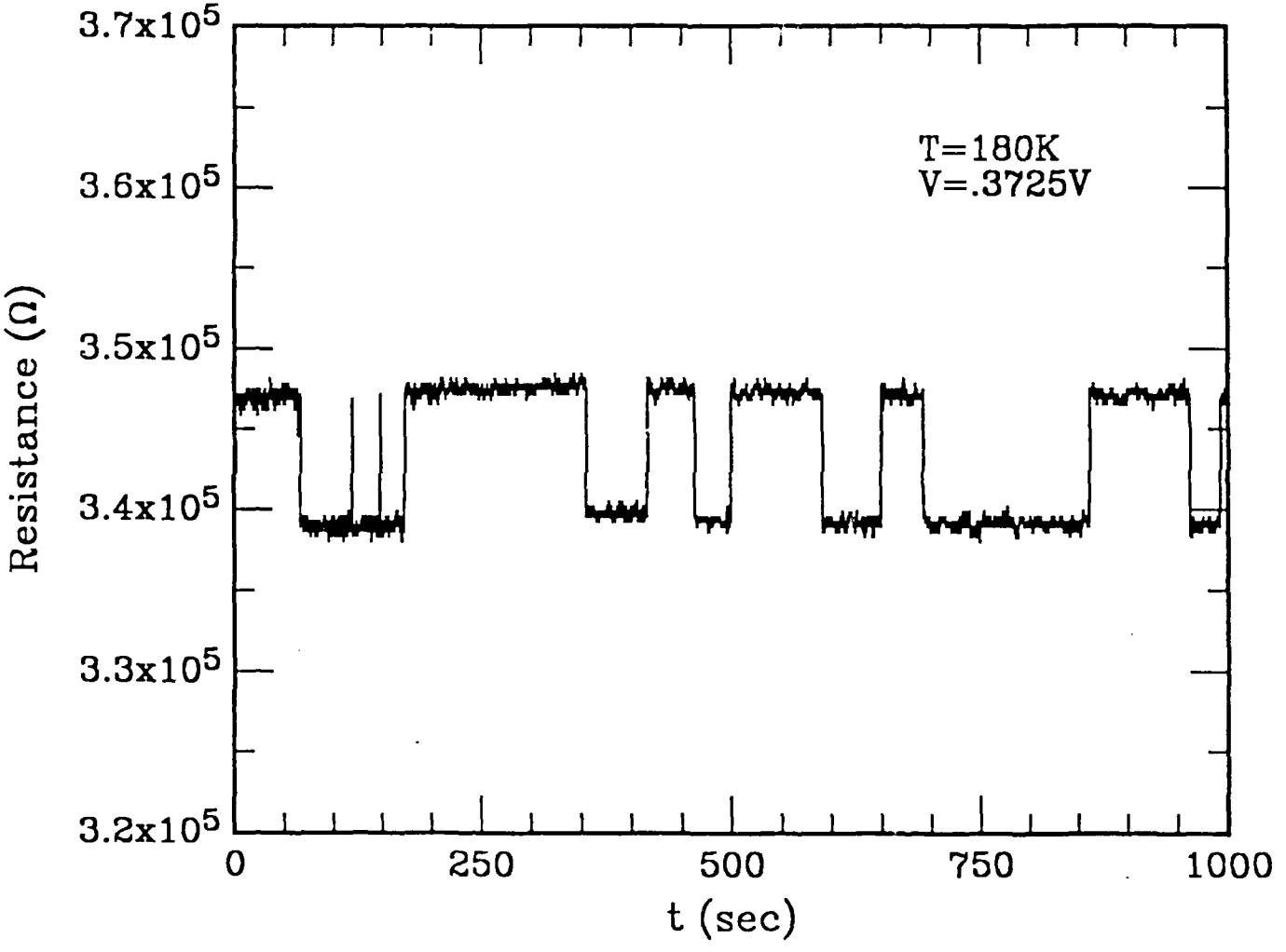

3.7x10^5
3.6x10^5
3.5x10^5
3.4x10^5
3.3x10^5
3.2x10^5
Resistance (Ω)
T=180K
V=.3725V
0
250
500
750
1000
t (sec)

FIG 10

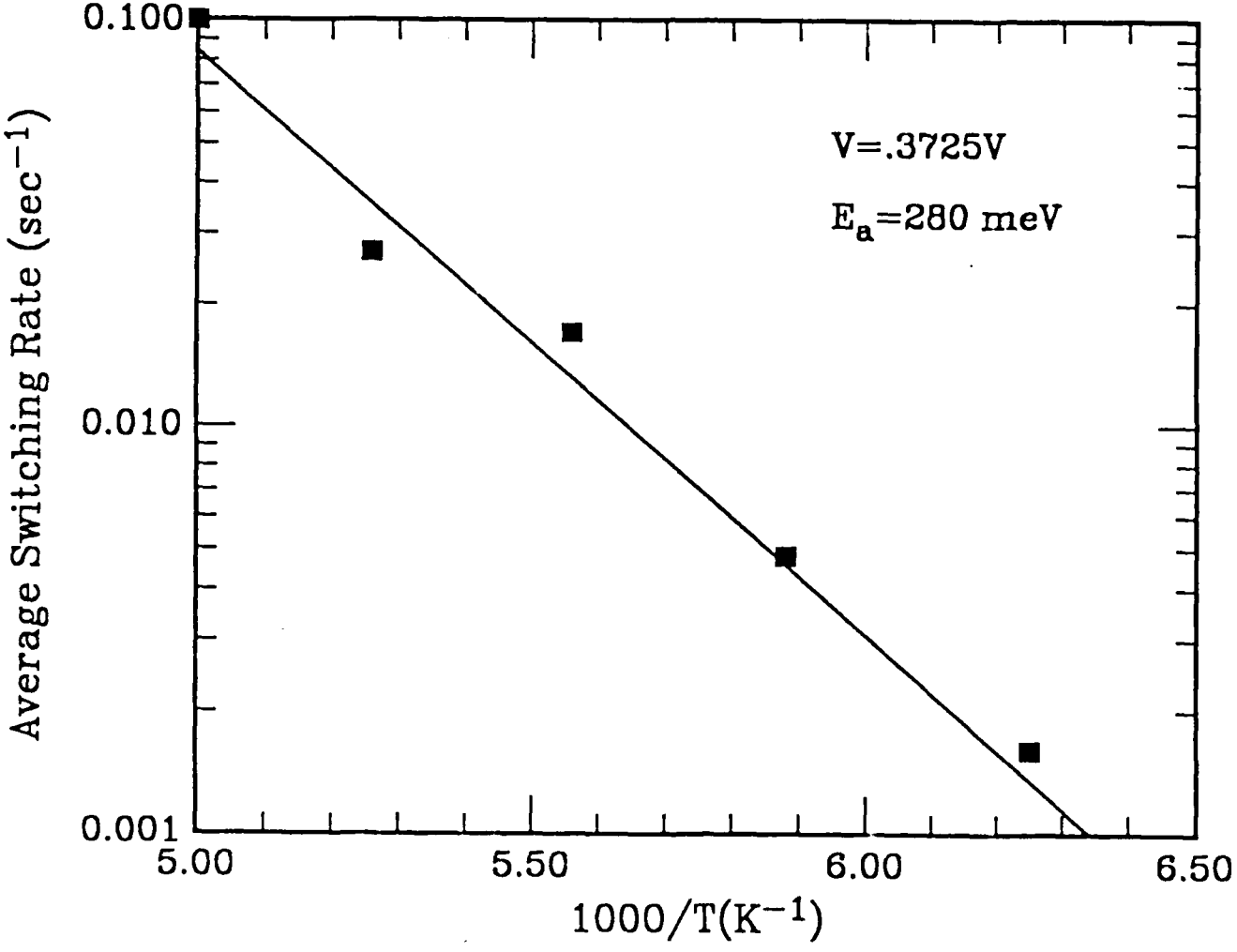

0.100
0.010
0.001
Average Switching Rate (sec^{-1})
V=.3725V
E_a=280 meV
5.00
5.50
6.00
6.50
1000/T(K^{-1})

FIG. 11

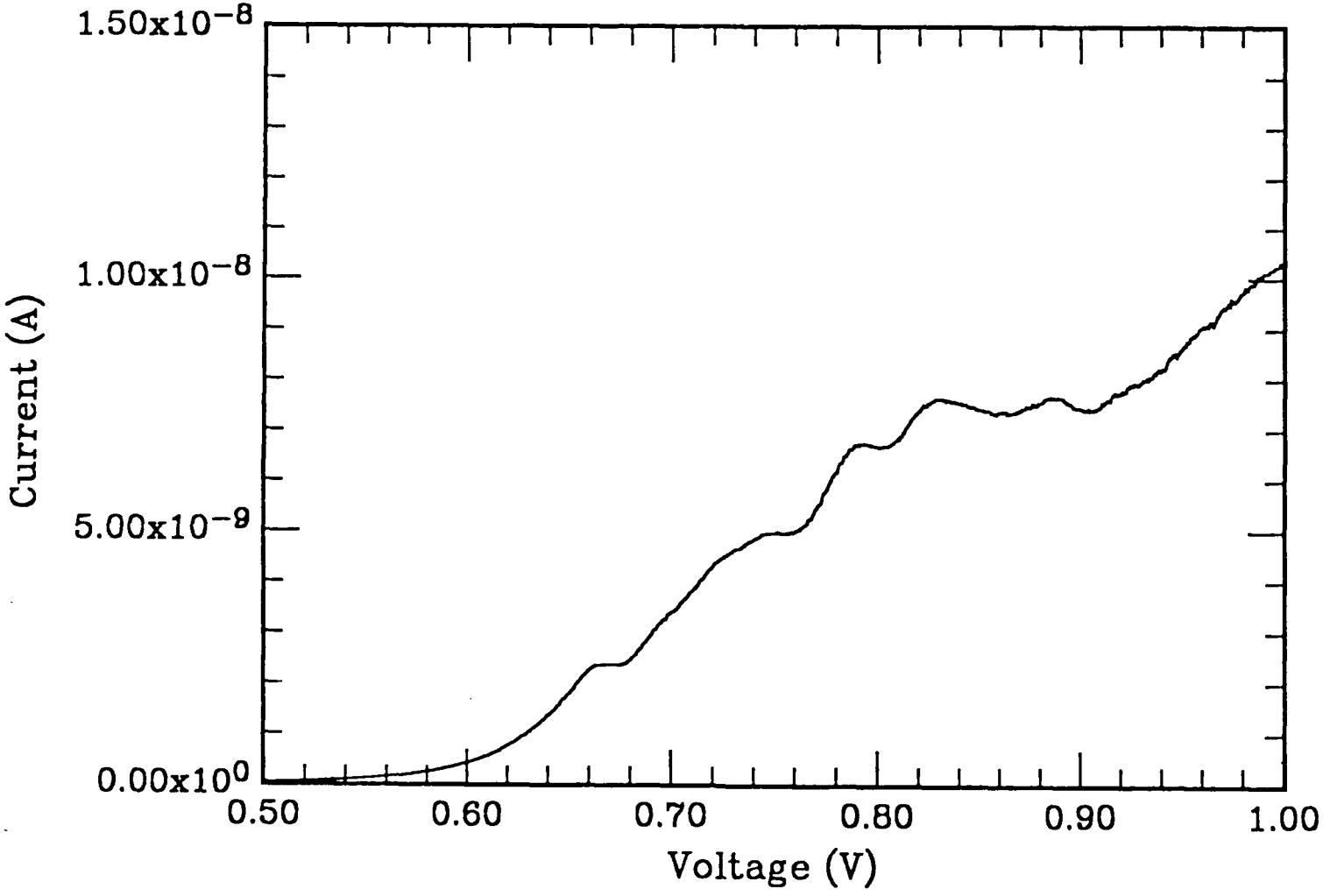

1.50x10^{-8}
1.00x10^{-8}
5.00x10^{-9}
0.00x10^{0}
Current (A)
0.50
0.60
0.70
0.80
0.90
1.00
Voltage (V)

FIG. 12

Appendix F

"Observation of Discrete Electronic States in a Zero-Dimensional Semiconductor Nanostructure"

[Published in Physical Review Letters 60, 535 (1988)].

Observation of Discrete Electronic States in a Zero-Dimensional Semiconductor Nanostructure

M. A. Reed, J. N. Randall, R. J. Aggarwal,[a] R. J. Matyi, T. M. Moore, and A. E. Wetsel[b]
Central Research Laboratories, Texas Instruments Incorporated, Dallas, Texas 75265
(Received 2 October 1987)

Electronic transport through a three-dimensionally confined semiconductor quantum well ("quantum dot") has been investigated. Fine structure observed in resonant tunneling through the quantum dot corresponds to the discrete density of states of a zero-dimensional system.

PACS numbers: 73.20.Dx, 72.15.Rn, 72.70.+m, 73.40.Gk

Carrier confinement to reduced dimensions in a semiconductor was first demonstrated in GaAs-AlGaAs quantum wells by electronic[1] and optical[2] spectroscopy in 1974. This achievement had led to numerous important developments in basic semiconductor physics and device technology. Structures produced by ultrathin-film growth are inherently two dimensional, and thus investigations have been largely confined to heterostructures where only the carrier momentum normal to the interfaces is quantized. Recent advances in microfabrication technology[3-5] have allowed the fabrication of structures with quantum confinement to one dimension ("quantum wires")[6,7] and have initiated intriguing investigations into one-dimensional physics, such as localization and electron-electron interaction,[8,9] single-electron trapping,[10] and universal conductance fluctuations.[11] It is expected that the realization of semiconductor heterostructures with quantum confinement to zero dimensions ("quantum dots") will yield equally intriguing phenomena. Attempts to observe confinement optically have been reported recently,[12-16] but the spectra do not show the characteristic structure of a series of isolated peaks expected from a zero-dimensional electron-hole gas. We have therefore studied such structures by electronic transport, and in this Letter present evidence for electronic transport through a discrete spectrum of states in a nanostructure confined in all three spatial dimensions.

The approach used to produce quantum-dot nanostructures suitable for electronic transport studies was to confine resonant-tunneling heterostructures laterally with a fabrication-imposed potential.[17] This approach embeds a quasibound quantum dot between two quantum-wire contacts. The initial molecular-beam-epitaxial structure is a 0.5-μm n^+-GaAs contact (Si doped at 2×10^{18} cm^{-3}, graded to approximately 10^{16} cm^{-3} over 200 Å, followed by a 100-Å undoped GaAs spacer layer), a 40-Å $Al_{0.25}Ga_{0.75}As$ tunnel barrier, and a 50-Å undoped $In_xGa_{1-x}As$ quantum well. The structure was grown to be nominally symmetric about a plane through the center of the quantum well. Employing a $In_xGa_{1-x}As$ quantum well allows one to lower the quantum well states with respect to the conduction-band edge while keeping the vertical dimensions fixed; x values studied ranged from 0 to 0.08. Large-area ($\geqq 2$ μm square) mesas of a typical structure ($x=0.08$) fabricated by conventional means exhibited two resonant peaks: a ground state at 50 mV with a peak current density of 30 A/cm^2, and an excited state at 700 mV with a peak current density of 8.1×10^3 A/cm^2, both measured at 77 K.

Electron-beam lithography defined an ensemble of AuGe/Ni/Au Ohmic metallization dots (single- or multiple-dot regions), nominally 1000–2500 Å in diameter, on the top n^+-GaAs contact by use of a bilayer polymethylmethacrylate (PMMA) resist and liftoff. The metal-dot Ohmic contact served as an etch mask for highly anisotropic reactive-ion etching with BCl_3 as an etch gas, defining columns in the epitaxial structure. A scanning electron micrograph of a collection of these etched structures is seen in Fig. 1. To make contact to the tops of the columns, a planarizing and insulating polyimide was spun on the sample and then etched back by O_2 reactive-ion etching to expose the metal contacts on the tops of the columns. A gold contact pad was then

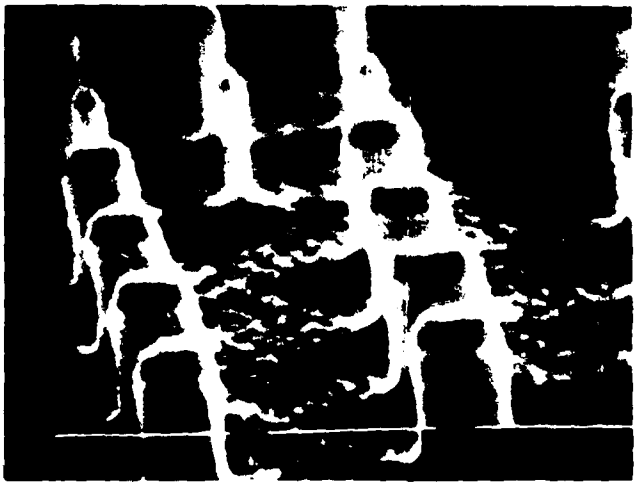


FIG. 1. A scanning electron micrograph of various size GaAs nanostructures containing quantum dots. The dark region on top of the column is the electron-beam defined Ohmic contact and etch mask. The horizontal bars are 0.5 μm.

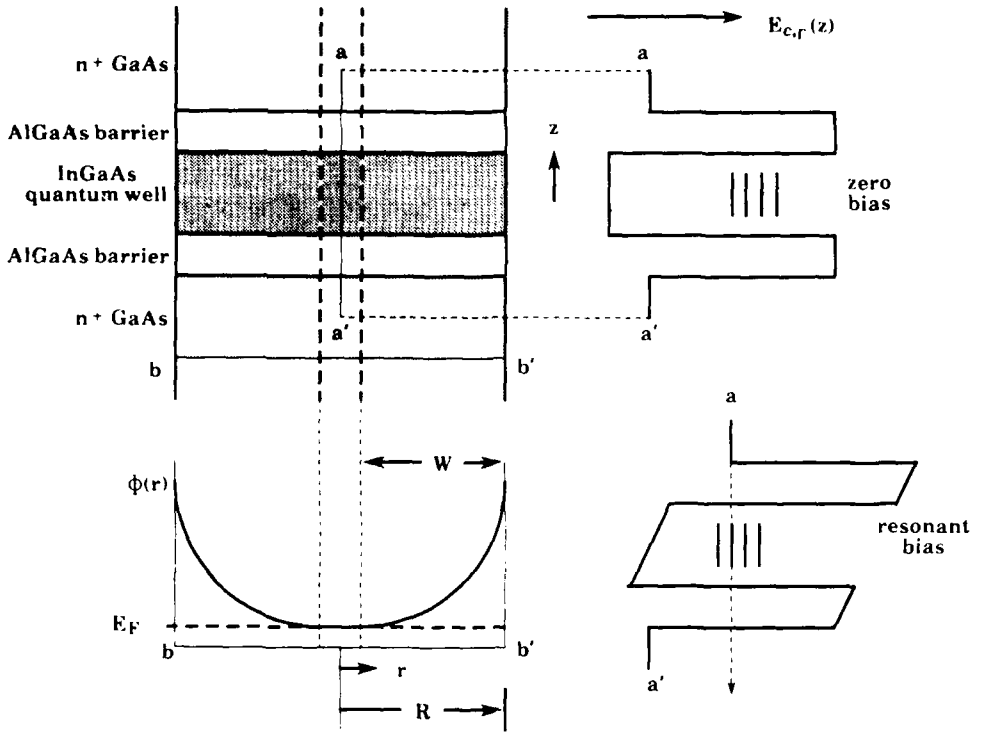


FIG. 2. Schematic illustration of the vertical (a-a') and lateral (b-b') potentials of a column containing a quantum dot, under zero and applied bias. $\Phi(r)$ is the (radial) potential, R is the physical radius of the column, r is the radial coordinate, W is the depletion depth, Φ_T is the height of the potential determined by the Fermi-level (E_F) pinning, and $E_{c,\Gamma}$ is the Γ-point conduction-band energy.

evaporated over the top(s) of the column(s). The bottom conductive substrate provided electrical continuity.

Figure 2 schematically illustrates the lateral (radial) potential of a column containing a quantum dot, and the spectrum of three-dimensionally confined electron states under zero and applied bias. A spectrum of discrete states will give rise to a series of resonances in transmitted current as each state drops below the conduction-band edge of the injection contact. To observe lateral quantization of quantum well state(s), the physical size of the structure must be sufficiently small that quantization of the lateral momenta produces energy splittings $> kT$. Concurrently, the lateral dimensions of the structure must be large enough that pinchoff of the column by the depletion layers formed on the sidewalls of the GaAs column does not occur. As a result of the Fermi-level pinning of the exposed GaAs surface, the conduction band bends upward (with respect to the Fermi level), and where it intersects the Fermi level determines in real space the edge of the central conduction-path core. We can express the radial potential $\Phi(r)$ in the column [for $(R-W) \leq r \leq R$], assumed axially symmetric, as

$$\Phi(r) = \Phi_T[1-(R-r)/W]^2, \quad (1)$$

where r is the radial coordinate, R is the physical radius of the column, W is the depletion depth, and Φ_T is the height of the potential determined by the Fermi-level pinning. When the lateral dimension is reduced to $2W$ or less, the lateral potential becomes parabolic though conduction through the central conduction-path core is pinched off.

A structure that satisfies both constraints was achieved with a $In_{0.08}Ga_{0.92}As$ quantum-well double-barrier structure with a physical (lithographic) lateral dimension of $\simeq 1000$ Å. Figure 3 shows the current-voltage characteristics of this (single) microstructure as a function of temperature. If we assume that the current density through the structure is approximately the same as in a large-area device, measurement of the peak resonant current implies a minimum (circular) conduction-path core of 130 Å for this structure; thus, a lateral parabolic potential approximation is valid. This implies a depletion depth of $\simeq 430$ Å at the double-barrier structure, in reasonable agreement with that expected from the known doping level (at 2×10^{18} cm^{-3}, $W = 220$ Å) and with the realization that W will enlarge in the undoped double-barrier region. The splitting of the discrete electron levels in the quantum dot is then

$$\Delta E = (2\Phi_T/m^*)^{1/2}\hbar/R, \quad (2)$$

where m^* is the effective mass of the electrons in the quantum well (linearly extrapolated between that of GaAs and InAs) and R is the physical radius. With a Fermi-level pinning of 0.7 eV, the states should be split

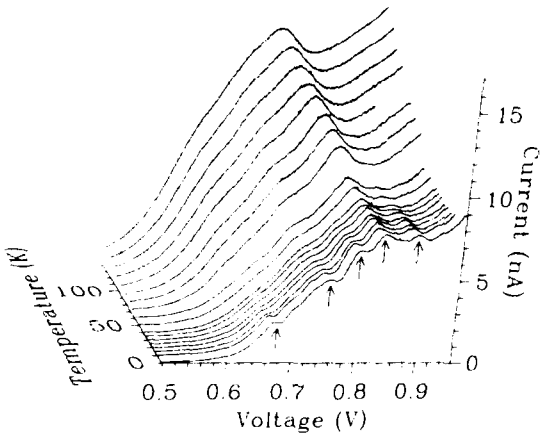


FIG. 3. Current-voltage characteristics of a single quantum-dot nanostructure as a function of temperature, showing resonant tunneling through the discrete states of the $n=2$ quantum well resonance. The arrows indicate the voltage positions of the discrete states for the $T=1.0$-K curve.

evenly by 26 meV.

Only the excited-state resonance of this structure is observed since the current expected from the ground-state resonance is at the detection limits of the test apparatus. At high temperature, the characteristic negative differential resistance of the double-barrier structure is evident. As the temperature is lowered, two effects occur. First, the resonant peak shifts slightly higher in voltage (because of the increase of the GaAs contact resistance with temperature) and decreases in current (because of the freezing out of excess leakage current). Secondly, there appears a series of peaks superimposed on the negative-differential-resistance peak. In the range 0.75–0.9 V the peaks are approximately equally spaced, with a splitting of $\simeq 50$ mV. Under the assumption that most of the bias is incrementally dropped across the double-barrier structure, the splitting of the equally spaced series is 25 meV, in excellent agreement with the value determined from the physical dimension of the structure. We believe that the structure observed here corresponds to resonant tunneling through the spectra of discrete quasibound (in the z direction) states in the quantum dot which correspond to the density of states of a three-dimensional semiconductor quantum well.

Another peak, presumably the ground state of the harmonic-oscillator potential, occurs $\simeq 80$ mV below the equally spaced series. The origin of this anomalously large splitting is not understood, though nonparabolicity of the lateral confining potential cannot be ruled out.[18]

In conclusion, we have measured electronic transport through quantum well states that have been laterally confined in all three dimensions (by heterojunction barriers in one dimension and a fabrication-imposed potential in the lateral dimensions) on the nanometer scale. We have performed electrical spectroscopy in the form of resonant tunneling through the spectrum of electron states, and observe resonances that correspond to the density of states of a zero-dimensional system.

We are indebted to R. T. Bate, W. R. Frensley, J. H. Luscombe, and R. H. Silsbee for helpful discussions and analysis, and to R. K. Aldert, D. A. Schultz, P. F. Stickney, and J. R. Thomason for technical assistance. This research was supported in part by the U.S. Army Research Office and the U.S. Office of Naval Research.

[a]Present address: Massachusetts Institute of Technology, Cambridge, MA 02139.

[b]Present address: Harvard University, Cambridge, MA 02138.

[1]L. L. Chang, L. Esaki, and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).

[2]R. Dingle, A. C. Gossard, and W. Wiegmann, Phys. Rev. Lett. **33**, 827 (1974).

[3]A. N. Broer, W. W. Molzen, J. J. Cuomo, and N. D. Wittels, Appl. Phys. Lett. **29**, 596 (1976).

[4]R. E. Howard, P. F. Liao, W. J. Skocpol, L. D. Jackel, and H. G. Craighead, Science **221**, 117 (1983).

[5]H. G. Craighead, J. Appl. Phys. **55**, 4430 (1984).

[6]H. Sakaki, Jpn. J. Appl. Phys. **19**, L735 (1980).

[7]W. J. Skocpol, L. D. Jackel, R. E. Howard, E. L. Hu, and L. A. Fetter, Physica (Utrecht) **117&118**, 667 (1983).

[8]R. G. Wheeler, K. K. Choi, A. Goel, R. Wisnieff, and D. E. Prober, Phys. Rev. Lett. **49**, 1674 (1982).

[9]T. J. Thornton, M. Pepper, H. Ahmed, D. Andrews, and G. J. Davies, Phys. Rev. Lett. **56**, 1198 (1986).

[10]K. S. Raals, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, Phys. Rev. Lett. **52**, 228 (1984).

[11]W. J. Sckocpol, P. M. Mankiewich, R. E. Howard, L. D. Jackel, D. M. Tennant, and A. D. Stone, Phys. Rev. Lett. **56**, 2865 (1986).

[12]M. A. Reed, R. T. Bate, K. Bradshaw, W. M. Duncan, W. R. Frensley, J. W. Lee, and H.-D. Shih, J. Vac. Sci. Technol. B **4**, 358 (1986).

[13]K. Kash, A. Scherer, J. M. Worlock, H. G. Craighead, and M. C. Tamargo, Appl. Phys. Lett. **49**, 1043 (1986).

[14]J. Cibert, P. M. Petroff, G. J. Dolan, S. J. Pearton, A. C. Gossard, and J. H. English, Appl. Phys. Lett. **49**, 1275 (1986).

[15]H. Temkin, G. J. Dolan, M. B. Panish, and S. N. G. Chu, Appl. Phys. Lett. **50**, 413 (1987).

[16]R. L. Kubena, R. J. Joyce, J. W. Ward, H. L. Garvin, F. P. Stratton, and R. G. Brault, Appl. Phys. Lett. **50**, 1589 (1987).

[17]J. N. Randall, M. A. Reed, T. M. Moore, R. J. Matyi, and J. W. Lee, to be published; A. E. Wetsel, M. A. Reed, J. N. Randall, R. J. Aggarwal, R. J. Matyi, and T. M. Moore, to be published.

[18]M. Luban and D. L. Pursey, Phys. Rev. D **33**, 431 (1986).

Appendix G

"Observation of Single Electron Trapping Phenomena
in Heterostructure Quantum Wires"

[Submitted to Physical Review B]

SUBMITTED TO
PHYS. REV. B

Observation of single electron trapping phenomena in heterostructure quantum wires

A. E. Wetsel,[a] M. A. Reed, J. N. Randall, R. J. Aggarwal,[b] R. J. Matyi and T. M. Moore

Central Research Laboratories

Texas Instruments Incorporated

P. O. Box 655936, MS 154

Dallas, Texas 75265

ABSTRACT

Discrete level resistance "telegraph noise" fluctuations have been observed in vertical GaAs microfabricated heterostructure quantum wires due to single electron trapping at defects. Several traps are observed and seem to work independently. The trapping rate, or the capture and escape of electrons from the defects, varies with temperature and is thermally activated. It is shown that the mechanism is fluctuation of the potential barrier in the vertical heterostructure. The change in inelastic tunneling due to scattering from a single trap is measured.

PACS numbers: 72.15.-v, 72.70.+m, 73.40.Gk, 73.40.Lq.

Advances in device fabrication have made it possible to study quantum effects in extremely small electronic devices. The effects of individual defects upon electronic transport through the device can be observed in detail. Recently, there has been a great deal of interest in telegraph noise exhibited in electronic microstructures such as metal-oxide-semiconductor field-effect transistors (MOSFET's)[1,2] and MOS tunnel diodes.[3,4] The discrete resistance fluctuations of the device are thought to be due to the trapping and escape of a single electron at a defect in a tunnel barrier or in an inversion layer. In this paper we report the observation of resistance fluctuations due to single trap states in microfabricated GaAs heterostructure quantum wires. Analysis of the trapping behavior in this structure allows us to determine if the effect is due to scattering by defects in the conduction channel or fluctuations of the barrier heights in the heterostructure. We have also observed a change in the amount of inelastic scattering due to a single one of these defects.

Large, discrete, two-level resistance fluctuations give clear information on a single trap affecting transport through a single quantum wire. In the case of MOSFET wires, as shown by Ralls *et. al.*,[2] electrons in the conduction channel are trapped or detrapped at defects, thus changing the charge state of the defect(s). The traps act as variable scattering centers in the quasi-1D conduction path of the MOSFET wires due to the difference in the scattering cross section of a neutral versus a charged defect. This causes a change in the impedance of the channel, which is manifested as discrete resistance changes

in the inversion layer. The capture and emission processes of the electrons are thermally activated, thus yielding the energetic position of the trap.

A conceptually different mechanism has been observed in MOS tunnel junctions,[4] in which trap states in the oxide layer fluctuate around the Fermi level until electron tunneling is favorable. The electrons tunnel through the top of the potential barrier, and as the electrons are trapped and detrapped at defects in the barrier the potential height of the barrier fluctuates. This modulates the tunneling current through the potential barrier, and thus the device impedance fluctuates. The conceptual distinction between the two cases is that, in the tunnel junction case the detailed potential in the region of the barrier is modified due to the trapping of an electron, whereas in the MOSFET wire embodiment the impedance is affected by the change of defect scattering cross sections in the conduction path.

To elucidate the differences between these two effects, we chose an embodiment of a vertical, microfabricated GaAs quantum wire that contains a tunnel junction;[5] specifically a double barrier/single quantum well resonant tunneling structure. The barriers serve as a current limiter and, as will be seen, allows us to distinguish between the two different mechanisms mentioned previously. This is a physical embodiment of a tunnel junction/quantum wire combination. The barriers in the device should give analogous results to the MOS tunnel diodes, and the narrow GaAs wire with the barrier in the conduction path should allow for scattering as in the narrow

MOSFET wires. Fabricating the structure this way gives the possibility of simultaneously observing and distinguishing between these mechanisms described above.

The resonant tunneling structures used in this study were grown by molecular beam epitaxy (MBE) in a Riber 2300 on a directly heated 2-inch n^+ (Si-doped, Sumitomo) GaAs substrate oriented two degrees off the $<100>$. The structure consists of two 50 Å undoped $Al_{.27}Ga_{.73}As$ barriers enclosing a 50 Å undoped GaAs quantum well. Two undoped 50 Å GaAs spacer layers were grown on either side of the barriers, with two 5000Å Si-doped GaAs buffer layers on top and bottom. The structure nominally has inversion symmetry about a plane through the center of the quantum well.

An e-beam lithography step and lift-off process followed, leaving a circular metal pad of ~2500Å diameter that served as an etch mask and top Ohmic contact. A highly anisotropic reactive ion etch (RIE) process etched down into the well structure forming columns (vertical wires) with the Ohmic contacts on top. To make contact with the tops of the wires, an insulating and planerizing polyimide layer was spun onto the sample. The tops of the columns were then uncovered by an O_2 RIE. Once the tops of the wires were revealed, a gold bonding pad was defined to provide electrical contact to the tops of the wires. A backside Ohmic contact provided electrical continuity.

The samples used in this study are single quantum wires of dimension 2500Å in diameter.

The current-voltage characteristics of one of these quantum wires at 100°K is shown in Figure 1. The discrete switching between two impedance states is observed at a number of different bias positions (0.6V, 0.75V-0.85V, 1.0V-1.1V, 1.25V-1.35V). These fluctuations are superimposed onto the well known negative differential resistance characteristic of the double barrier/single quantum well heterostructure (due to the quantum well state passing through the conduction band edge of the contact). The overall structure of the characteristics were the same upon repeated bias sweeps, though the detailed structure in the fluctuation regions were not; in particular, the characteristics in the non-fluctuating regions (e.g., 0.65V-0.75V, 0.9V-0.95V, etc.) were exactly repeatable, but the fluctuating regions were not repeatable in detail.

The rate at which the impedance of the device switches between discrete values is a strong function of temperature. Figure 2 shows the resistance of a similar device, at fixed bias, as a function of time for different temperatures. Since the resistance fluctuates only between two discrete values (qualitatively identical to what has been previously observed[2,3]), the behavior indicates the trapping of a single electron onto a given trap in the structure. It should be noted that amplitude of the fluctuations is nearly constant over the bias range in which the trapping occurs (e.g., 0.75V-0.85V in Figure 1), implying that a

single trap only is involved. The difference in the amplitudes for different bias regions implies that more than one trap exists that causes the fluctuations. These traps are selectable by device bias. Clearly, the details of the current-voltage characteristics of such a device (i.e., Figure 1) depends not only on the temperature but on the rate that the bias is changed.

To verify that the resistance fluctuations are due to trapping onto a single defect, the average rate at which the resistance fluctuates was determined (at fixed bias) as a function of temperature. This is shown in Figure 3. The behavior is clearly activated, and measures the depth of the trap to be 280meV. Measurement on other traps in this (and other similar) device(s) indicated a wide distribution of trap energies, with switching due to some traps observable to room temperature.

In some instances the simultaneous effects of two different traps are observed. Figure 4 shows an example of a resistance increase to an "up" state (ΔR_1) and larger decrease to a "down" state (ΔR_2) due to different traps. These traps appear to act independently, since the impedance when both traps "switch on" is simply the sum of the impedances (i.e., at $\tau \sim 100$ sec, $\Delta R = \Delta R_1 - \Delta R_2$). These two traps clearly have different activation energies, and appear to be separated sufficiently well in real space to eliminate interaction effects.

The physical lateral size of the quantum wire appears to be large in comparison to the size of systems in which telegraph noise has previously been observed. However, this system is distinct from the other above mentioned cases due to the Fermi level pinning of free GaAs surfaces and the accompanying depletion layer that will extend laterally into the wire. Large ($\geq$2μm diameter) mesas fabricated by conventional means yielded a current density at the resonance peak of 1.6×10^4 A/cm^2. Assuming that the trap switching phenomena are a perturbation, we arrive at an effective (circular) conduction path diameter of ~500Å from the peak resonance current of Figure 1, consistent with the physical size scale necessary to observe telegraph noise in MOSFET wires or MOS tunnel junctions.

The negative differential resistance characteristic of the heterostructure quantum wire enables us to determine which mechanism (scattering or potential fluctuation) is causing the observed switching phenomena. Let us first consider the case of scattering. If the defect in the wire is a positively charged defect, capture of an electron converts it to a neutral, thus causing a decrease in the wire impedance. This will cause an increase in the current through the device. Thus, as the trap is biased into occupancy, the current state will switch from a low to asymptotically high current state. Let us define this sense of asymptotic current value "switching high". Similarly, a neutral trap capturing an electron to become a negatively charged trap will cause a decrease in the current, which asymptotically "switches low" with

occupancy. The situation of a neutral going into a negative is schematically shown in Figure 5(a).

As we have seen in Figure 1, these traps become occupied at various bias positions. The scattering mechanism predicts that the sense of switching ("high" or "low") will be the same no matter where in bias position the trap becomes occupied. Thus, the sense of switching will be the same on either side of the resonant peak position; i.e., "symmetric".

The second case to consider is a fluctuation in the local potential of the double barrier structure. The change in the charge state of a defect located in the accumulation, double barrier, or depletion regions of the structure can modulate the potential distribution across the double barrier structure. This shift in potential will move the position of the resonant tunneling state with respect to the conduction band edge of the injection contact, thus changing the tunneling current through the structure. For example, if the device is biased such that the quantum well state is resonant with the conduction band edge of the contact, and a trapping event causes the quantum well state to shift up in energy with respect to the conduction band edge, a decrease in current will occur. This can be thought of as an effective shift in the local current-voltage characteristics. Capture of an electron by a positively charged defect to become neutral, or of a neutral defect to become a negative, causes an effective shift to higher biases of the local current-voltage characteristics.

In this case, the asymptotic sense of current direction depends on where in bias position the trap becomes occupied. This is schematically illustrated in Figure 5(b). At device biases below the resonant peak position, the asymptotic sense is "switching low" as the trap becomes occupied; in the negative differential resistance region, the sense is "switching high"; and at large biases, the sense will eventually be "switching low". The sense of switching is not the same on either side of the resonant peak position; i.e., this is an "asymmetric" case.

As can be seen in Figure 1, this device exhibits asymmetry in the sense of switching. In the region 0.75V-0.85V, the current switches "low" as the trap becomes occupied. In the region 1.0V-1.1V, the current is switchs "high", and at very high bias (1.3V-1.35V) the state switches "low". Note that the structure in the range 1.2V-1.275V appears to first switch high, then low. This is consistent with the intersection of the shifted current-voltage characteristics occurring at approximately 1.26V.

The change in the charge state of these traps can not only shift the local potential but can cause changes in the dynamics of the resonantly tunneling electrons. Figure 6(a) show the current-voltage characteristic of a similar device that has a relatively flat resonance peak due to excess inelastic tunneling. Switching phenomena is observable for a trap that fortuitously becomes occupied at the resonant peak, and the impedance fluctuations due to

this trap exhibit marked modulation as the quantum well resonant state is passed through resonance. This cannot be explained by a simple effective shift in bias; indeed, the amplitude fluctuations are largest for minimum slope, opposite to what would be expected from a bias shift for this flat resonance. However, the observation is consistent with the switching between a given characteristic and a shifted characteristic with less inelastic scattering due to the neutralization of a positively charged defect. Figure 6(b) schematically illustrates this phenomena, assuming less inelastic scattering due to the charged defect becomming neutral. The amplitude of the switching first grows, than decreases as the device is biased through resonance. We are assuming a relatively small shift in the local potential due to this trap.

By observing the change in the amplitude fluctuations as the device is biased through resonance, we can determine the amount of inelastically scattered current (at resonance) due to a single trap. Assuming that neutral impurity scattering is negligible, we arrive at a value of 5.9×10^{-9} A/cm^2 per defect for inelastic scattering from positively charged defects, at resonance. We are thus able to uniquely measure the inelastic scattering cross section of a defect in a tunneling structure, which is equal to 3.7×10^{-13} cm^2 in this embodiment.

In conclusion, we have observed discrete resistance changes due to trap defects in GaAs heterostructure quantum wires. These traps are thermally activated and appear to act independently of each other. We have determined

that the mechanism responsible for the switching is the modification of the local potential in these structures due to trapping onto defects in the structure. We have also measured the inelastic scattering cross section of a single charged defect in vertical electronic transport through a heterostructure.

We would like to thank W. R. Frensley for helpful discussions and R. Aldert, D. Schultz, P. Stickney and R. Thomason for technical assistance. This research was supported in part by the Army Research Office and the Office of Naval Research.

REFERENCES

(a) Present address: Harvard University, Cambridge, MA.

(b) Present address: Massachusetts Institute of Technology, Cambridge, MA.

1 W. J. Skocpol, L. D. Jackel, E. L. Hu, R. E. Howard and L. A. Fetter, *Phys. Rev. Lett.* **49**, 951 (1982).

2 K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth and D. M. Tennant., *Phys. Rev. Lett.* **52**, 228 (1984).

3 C. T. Rogers and R. A. Buhrman, *Phys. Rev. Lett.* **55**, 859, (1985).

4 K. R. Farmer, C. T. Rogers, and R. A. Buhrman, *Phys. Rev. Lett.* **58**, 225 (1987).

5 J. N. Randall, M. A. Reed, T. M. Moore, R. J. Matyi, and J. W. Lee, Proceedings of the 31st International Symposium on Electron, Ion, and Photon Beams, May 1987, Boston (to be published in *J. Vac. Sci. Technol.*)

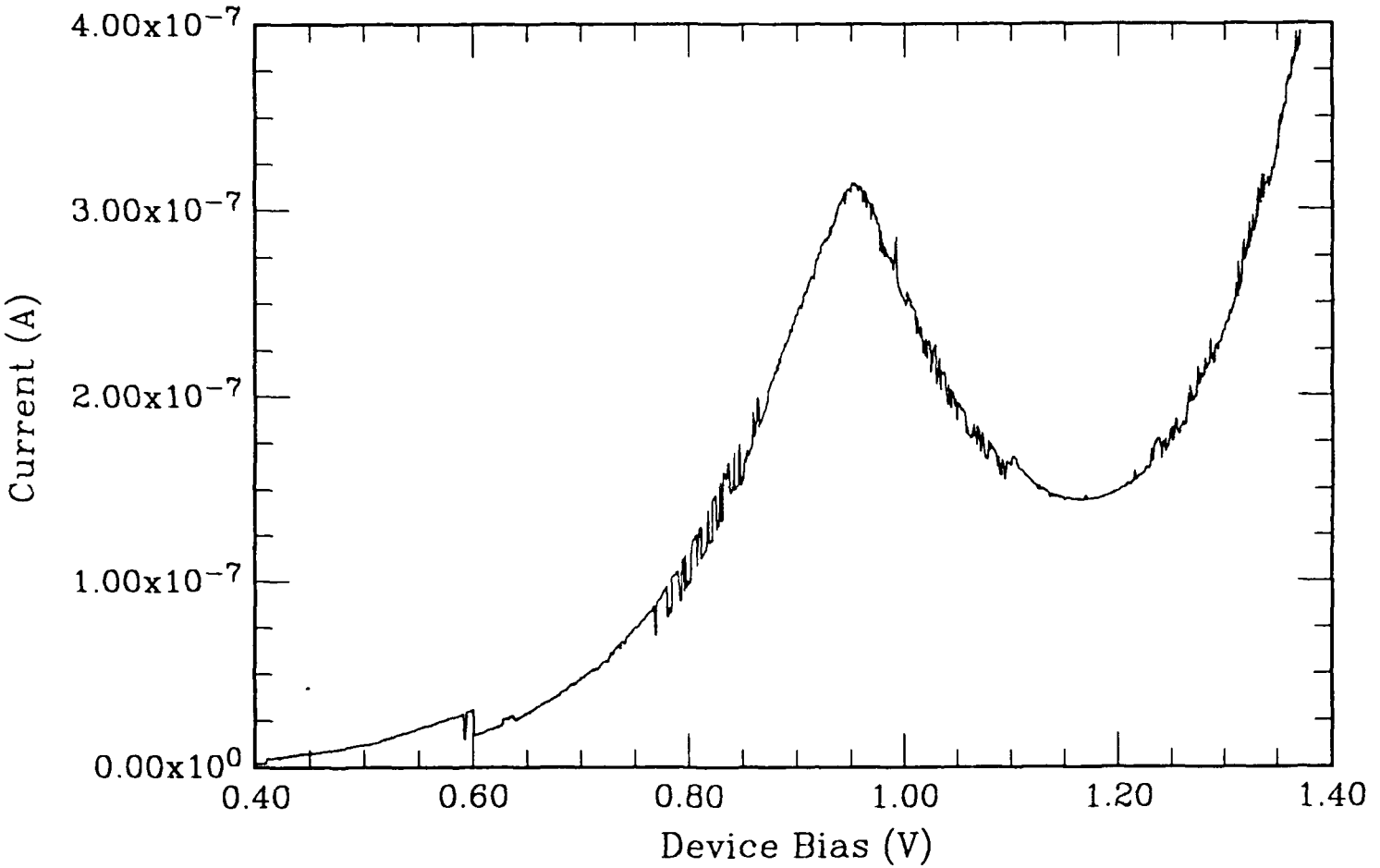
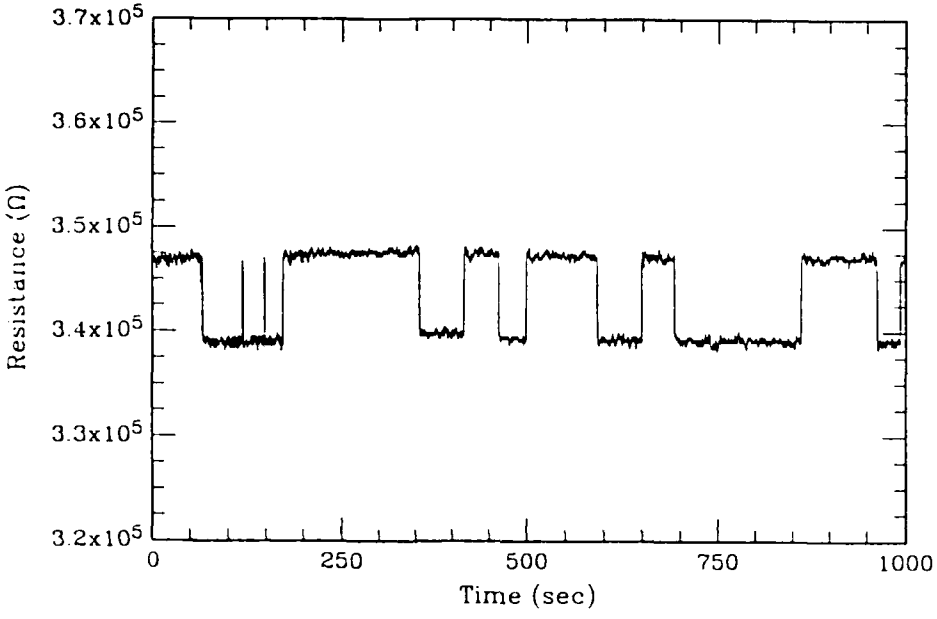
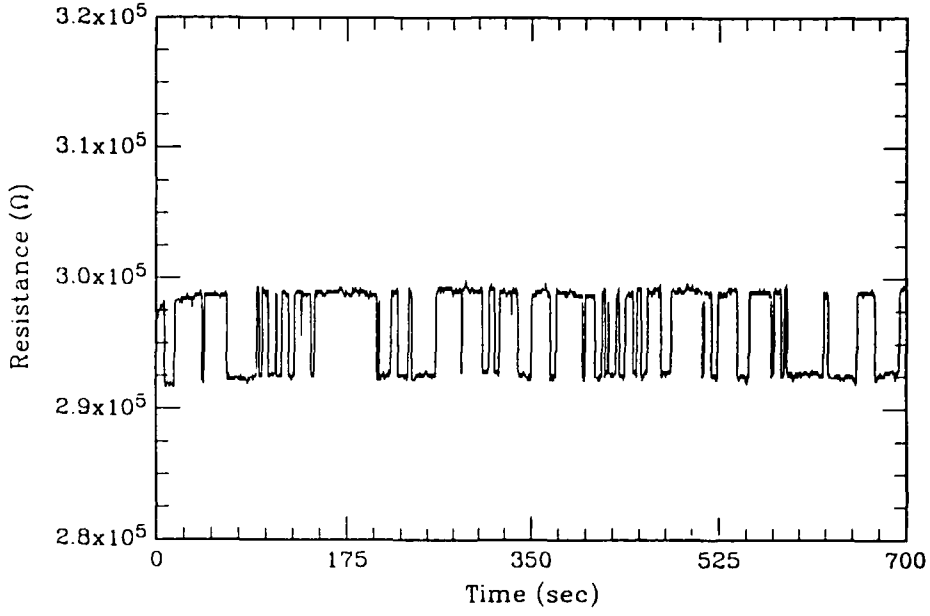


Figure 1. Current-voltage characteristic of a single heterostructure quantum wire at 100°K. The physical diameter of the wire is approximately 2500Å.



(a)



(b)

Figure 2. Resistance as a function of time for a single heterostructure quantum wire, at fixed bias potential, for a) T=180K, and b) T=200K.

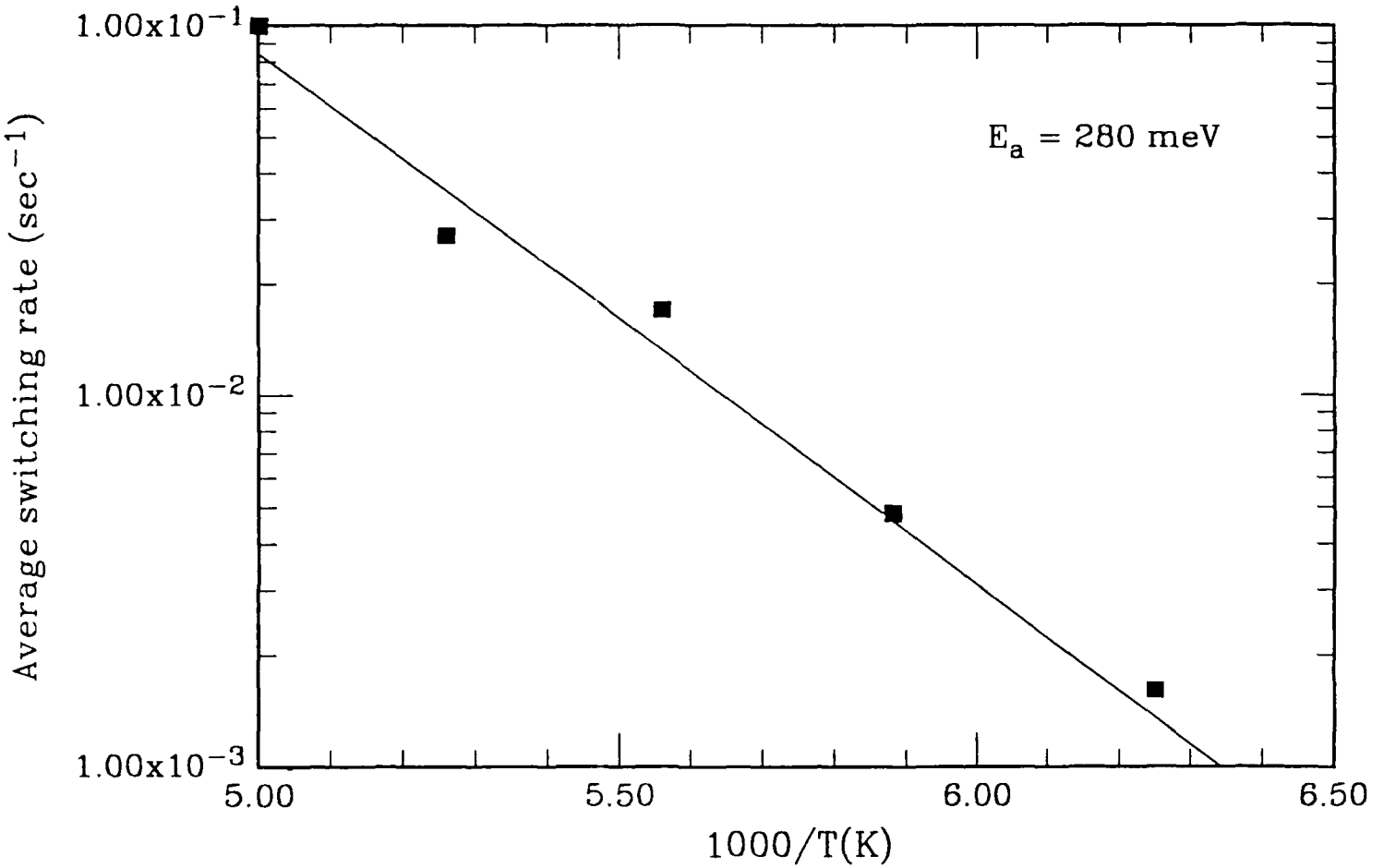


Figure 3. Average switching rate of a trap in a heterostructure quantum wire, at fixed bias. The straight line yields an activation energy of 280meV.

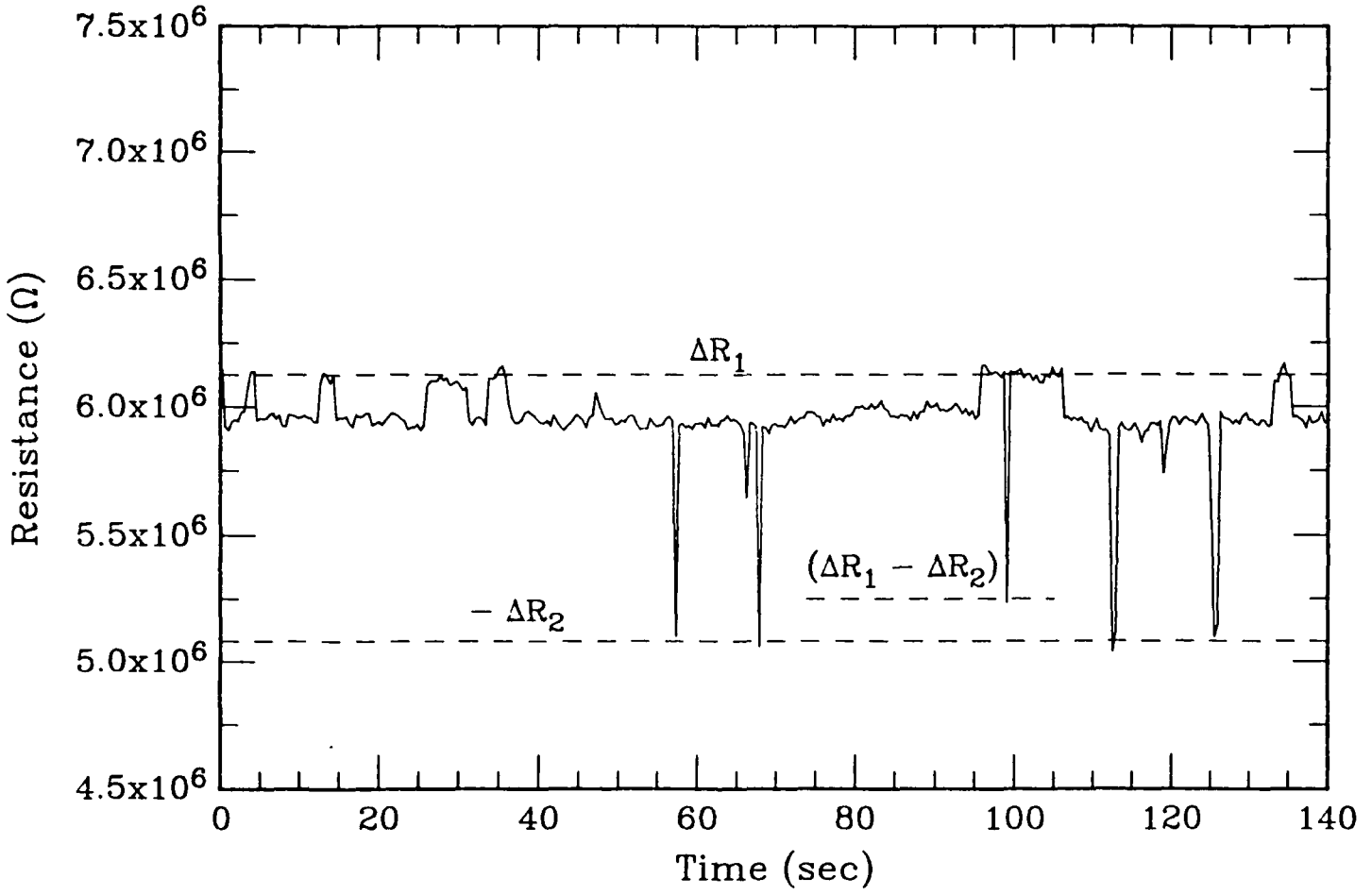
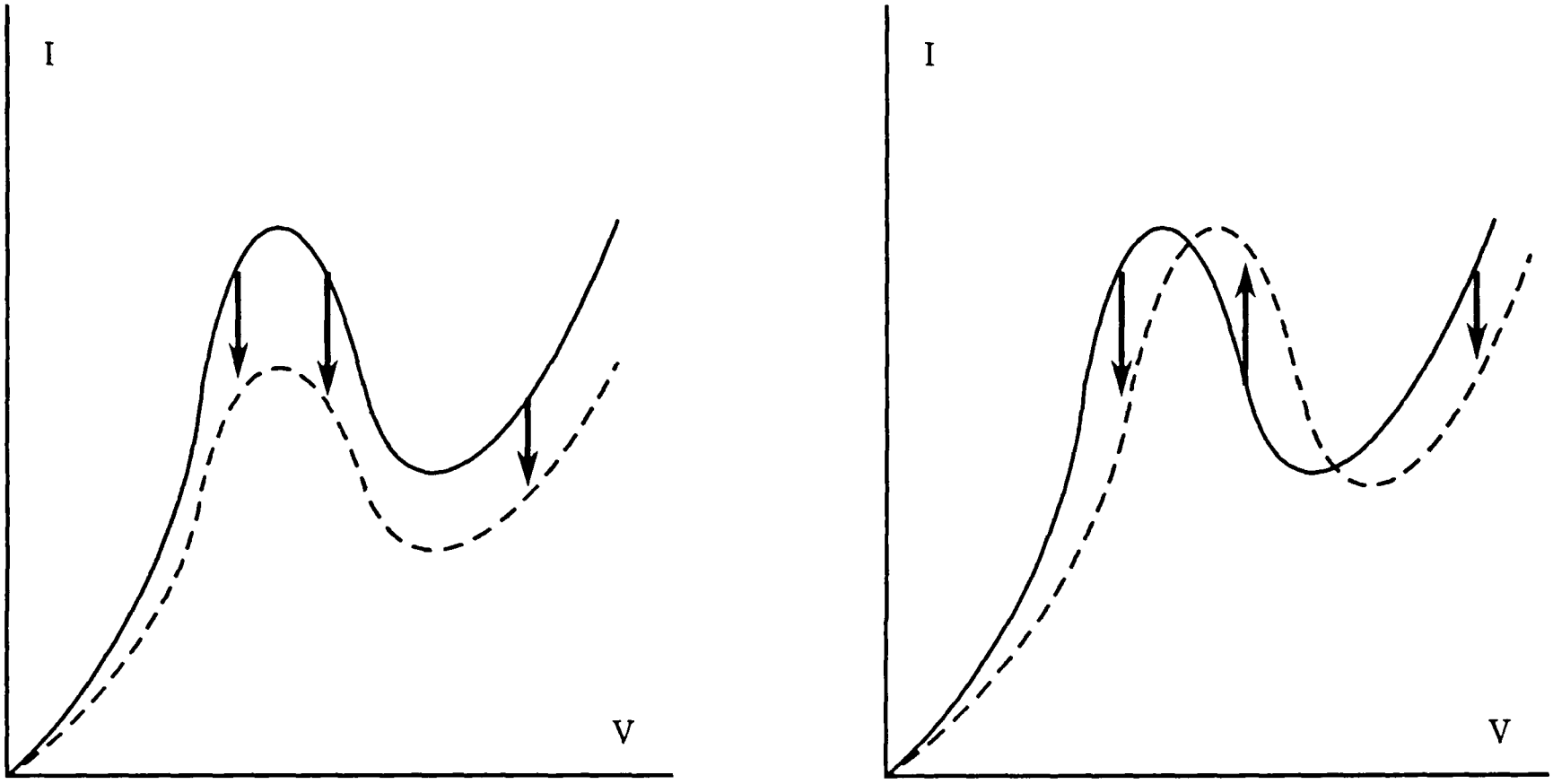
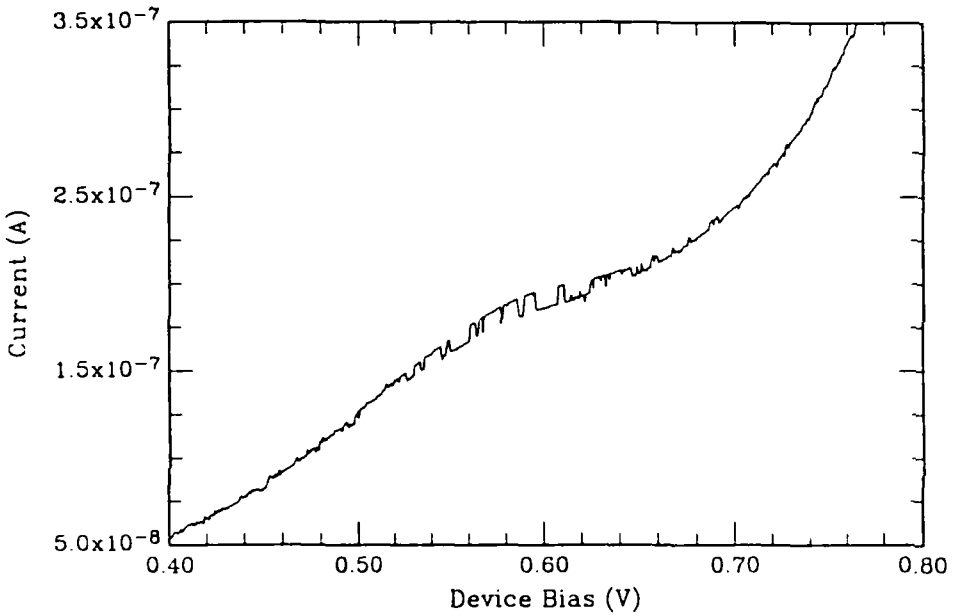


Figure 4. Resistance as a function of time for a single heterostructure quantum wire. The effects of two independent traps (ΔR_1 and ΔR_2) and simultaneous switching due to both (ΔR_1-ΔR_2) is observed.

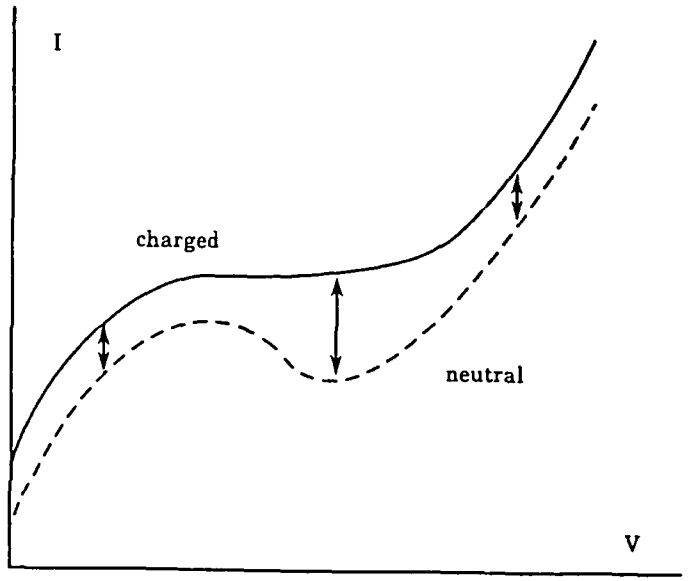


(a) (b)

Figure 5. Schematic current-voltage characteristics illustrating the sense of switching due to (a) scattering, and (b) potential fluctuations. In (a), the shift to the dashed curve indicates trapping onto a neutral impurity, and in (b) the shift to the dashed curve indicates trapping onto either a neutral or charged impurity. The arrows indicate the asymptotic sense of switching.



(a)



(b)

Figure 6(a). Current-voltage characteristic of a heterostructure quantum wire exhibiting a modulation in the switching amplitude. (b).Schematic current-voltage characteristics showing a structure with a charged defect state (solid line), and one with a neutral defect state (dashed line) with reduced inelastic scattering. The arrows indicates the amplitude of the switching.

Appendix H

"Quantitative Resonant Tunneling Spectroscopy: Current-Voltage Characteristics of Precisely Characterized RTDs"

[Presented at the 15th International Symposium on GaAs and Related Compounds, Atlanta, September 1988; to be published in Applied Physics Letters]

Quantitative resonant tunneling spectroscopy: Current-voltage characteristics of precisely characterized RTDs

M. A. Reed, W. R. Frensley, W. M. Duncan, R. J. Matyi*,
A. C. Seabaugh, and H.-L. Tsai

Central Research Laboratories, Texas Instruments Incorporated
Dallas, Texas 75265

ABSTRACT

A systematic comparison of precisely characterized resonant tunneling structures is presented. A self-consistent bandbending calculation is used to model the experimentally observed resonant peak positions. It is found that the peak positions can be accurately modeled if the nominal characterization parameters are alloweed to vary within the measurement accuracy of the characterization. As a result, it is found that the asymmetries in the current-voltage characteristics are solely explainable by tunnel barrier thickness fluctuations.

* Present address: Department of Metallurgical and Mineral Engineering, University of Madison, Wisconsin.

The origin of negative differential resistance in double barrier/single quantum well resonant tunneling structures is qualitatively well understood.[1] However, a full accounting of the current-voltage characteristics requires precise physical and electrical measurement of the device material properties. We present here a systematic comparison of measurements on precisely characterized resonant tunneling structures with models of the current-voltage dependence.

A set of four specimens has been grown by MBE to provide devices which vary over seven orders of magnitude in resonant peak current density for the AlGaAs/GaAs/AlGaAs system. This is achieved by growing nominally identical structures which differ solely in barrier thickness. Photoluminescence test structures were grown to provide for measurement of the AlGaAs band gap, transmission electron microscopy was used to independently verify the layer thicknesses, and capacitance-voltage profiling provided an independent determination of the doping density. The current-voltage characteristics of these structures have been measured. A systematic shift of the resonant peak and a variation of the resonant peak voltage asymmetry with current density is observed and compared with modeling results. It is found that a self-consistent bandbending model can accurately predict the voltage peak positions, though the structural parameters used are not necessarily the "nominal" values, yet values within the error of the characterization measurement.

The samples used in this study were grown on Si-doped n^+ GaAs conductive substrates using a Riber 2300 MBE system. The structures consist of a 0.5 micron Si-doped GaAs buffer and bottom contact layer, a (nominally) 150Å undoped GaAs spacer layer, an undoped AlGaAs tunnel barrier, an undoped GaAs quantum well, a second undoped AlGaAs tunnel barrier of nominally identical thickness, another 150Å undoped GaAs spacer layer, a 0.5 micron Si-doped GaAs top contact, a 0.5 micron undoped AlGaAs layer, an undoped GaAs 50Å quantum well, a 0.1

micron undoped AlGaAs layer, and a 100Å GaAs cap layer. The entire structure was grown at constant temperature at 600C (to minimize Si diffusion) as measured by a short wavelength pyrometer. The four specimens were grown sequentially to insure a constant unintentional impurity background.

The active resonant tunneling structure is buried underneath a series of diagnostic photoluminescence structures. Conventional photoluminescence was performed at 4.2K and 300K. The top quantum well photoluminescence exhibited typical FWHM values of 10 meV. Photoluminscence of the thick AlGaAs layers was used for determination of the Al content. The band gap relation of Casey and Panish[2] was assumed. The doping density of the n+ GaAs layers was determined by standard capacitance-voltage profiling measurements, and the thicknesses of the tunnel barriers and quantum wells were determined by cross-sectional TEM. A summary of these structural parameters for the four samples studied is shown in Table I.

Prior to device fabrication, the top diagnostic layers were removed by a chemical etch so that contact could be made to the upper n+ GaAs layer. Mesa devices ranging from 1.6×10^{-7} cm^{-2} to 4.1×10^{-5} cm-2 were fabricated by standard photolithography and chemical etching. Bonding pads contacted the upper top AuGeNi alloyed metal Ohmic contact and a similar bottom contact through a Si_3N_4/polyimide passivation layer. Static current-voltage characteristics (4-point where necessary) were measured at 77K.

Figure 1 shows a realistic conduction energy band profile of the 85Å barrier thickness structure under (a) zero and (b) resonant bias. The model from which this Figure was obtained finds the self-consistent solution of Poisson's equations for the electrostatic potential. The electrons in the contacts are treated in a finite-temperature Thomas-Fermi approximation. (i.e. these electrons are assumed to be in local equilibrium with the Fermi levels established by their respective electrodes.)

One result of this calculation, illustrated in Figure 1(a), is that the band profile near the quantum well is significantly perturbed by the contact potential of the n+-undoped junction. This shifts the resonant state upward (with respect to the n+ GaAs Fermi level) from that expected from a naive flat-band picture. This contact potential thus shifts the resonant peak position (Figure 1(b)) considerably; the model predicts a resonant voltage at 310 meV, much higher than that predicted by a flat-band picture.

Figure 2 shows the experimental current-voltage characteristics of a typical (4 micron)^2 mesa device of this structure at 77K. Care must be exercised in the spectroscopy of the structures. Current-voltage characteristics of successively increasing mesa size (same epitaxial structure) progressively exhibits the well-known plateau structure due to self-biasing[3]. This self-biasing perturbs and is observed to lower the apparent resonant peak position. For accurate spectroscopy, this effect must be avoided.

The experimental resonant peaks are not in very good agreement with the model calculation; the experimental peaks appear at 263 meV and 227 meV for positive and negative bias polarity, respectively, whereas the model predicts a value of 310 meV. (The convention here is that positive bias polarity implies electron injection from the top epitaxial contact). What is also obvious is the asymmetry of the resonant peaks holds for both voltage and current. It is found that this asymmetry is not consistent for a large sampling of similar devices; the asymmetry ranges from zero to as much as 60 meV in voltage and a factor of 3.3 in current. However, the degree of asymmetry is correlated; a larger voltage asymmetry implies a larger current asymmetry.

To ascertain the degree of asymmetry and variation, characteristics of a large number of devices from the various epitaxial structures were measured. The resonant voltage peak positions as a function of resonant current density are shown

in Figure 3 (due to the above mentioned complication of stabilizing oscillations, measurements from the the 30Å barrier were unreliable and are not presented here). The 118Å barrier structure data exhibits a clear exponential behavior over two orders of magnitude, with the positive bias peaks occurring at lower voltages and current densities than the negative bias peaks. The 85Å barrier structure data is not as clear, exhibiting considerable scatter. The origin of this scatter is not known. Additionally, the data exhibits the inverse of the 118Å data; the positive bias peaks occur at higher voltages and current densities than the negative bias peaks. Finally, the 65Å barrier structure deviates significantly from exponential behavior.

Examination of the 118Å barrier structure data reveals the major cause of the asymmetry, both in current and voltage position. Consider a fluctuation in the thickness of one of the tunnel barriers of a nominal thickness barrier sample. This implies a change in the voltage position at which resonance occurs, and concurrently a change in the tunneling current. Figure 4 illustrates the 85Å barrier structure with parameters varied within the error bars quoted in Table I; specifically, with the top barrier thickness equal to 80Å, the bottom barrier thickness equal to 90Å, and the quantum well equal to 48Å. Figure 4(a) shows the modified structure under positive bias (the right hand side of the Figures corresponds to the top Ohmic contact), and exhibits a resonant voltage at 270 meV. In the reverse bias direction (Figure 5(b)), resonance occurs at 230 meV. These are in excellent agreement with experimentally observed values. Note that larger current densities correspond to electron injection first through the thinner (top) barrier.

Figure 3 shows that the inherently thinner top barriers of the 85Å sample are a sample-dependent phenomenon; the 118Å data reveal that the top barrier is thicker than the bottom barrier in the 118Å sample. Likewise, the 65Å sample appears to have approximately equal barrier thicknesses. These results imply that Si dopant redistribution, at least in these samples, is not a complication. Finally, the scatter in

the 85Å data with respect to the 118Å and 65Å data may imply that this sample has larger quantum well thickness fluctuations , though this cannot be verified without further data on Al content and doping fluctuations.

The quantitative spectroscopy of these structures is relatively straightforward if one stays in the regime where the structure impedence is dominated by the tunnel barriers. Outside of this regime (e.g., for the 65Å data) the device may be affected by an internal series resistance. The resonant voltage position for the 65Å data is found to be linear with current density, and gives a contact resistance of 8.8×10^{-5} Ω-cm^2, equal for both positive and negative bias peak positions. This resistance can be fully accounted for by the AuGeNi Ohmic metallization used here.

We have shown a self-consistent bandbending model that can accurately predict experimentally observed resonant peak positions, and have compared it with precisely characterized resonant tunneling structures. It is found that, to accurately model the resonant voltage peak positions, the characterization values must be varied within the error bars of the measurement. Indeed, this technique can be used as an accurate diagnostic of the structure. Asymmetries in the electrical characteristics have been shown to be due to fluctuations in the tunnel barrier thicknesses.

We are thankful to R. K. Aldert, R. T. Bate, J. N. Randall, P. F. Stickney, F. H. Stovall, J. R. Thomason, and C. H. Yang for discussions and technical assistance. This work has been supported by the Office of Naval Research.

REFERENCES

1. L.L. Chang, L. Esaki and R. Tsu *Appl. Phys. Lett.* **24** 593 (1974).
2. H.C. Casey and M.B. Panish *Heterostructure Lasers* (New York: Academic) pp 187-194 (1978).
3. J.F. Young, B.M. Wood, H.C. Liu, M. Buchanan, D. Landheer, A.J. SpringThorpe and P. Mandeville *Appl. Phys. Lett.*, **52** 1398 (1988).

TABLES

Table I

Barrier thickness (TEM)	RTD QW thickness (TEM)	Al content (PL, 300K)	PL QW energy (PL, 4.2K)	contact doping density, cm-3 (CV)
118(±5)Å	48(±5)Å	27.7(±0.6)	1.620 eV	$1.7(\pm 0.2)10^{18}$
85(±5)Å	44(±5)Å	26.4(±0.6)	1.623 eV	$1.7(\pm 0.2)10^{18}$
65(±5)Å	44(±5)Å	27.7(±0.6)	1.613 eV	$1.4(\pm 0.4)10^{18}$
32(±5)Å	38(±5)Å	25.0(±0.6)	1.616 eV	$2.6(\pm 0.1)10^{18}$

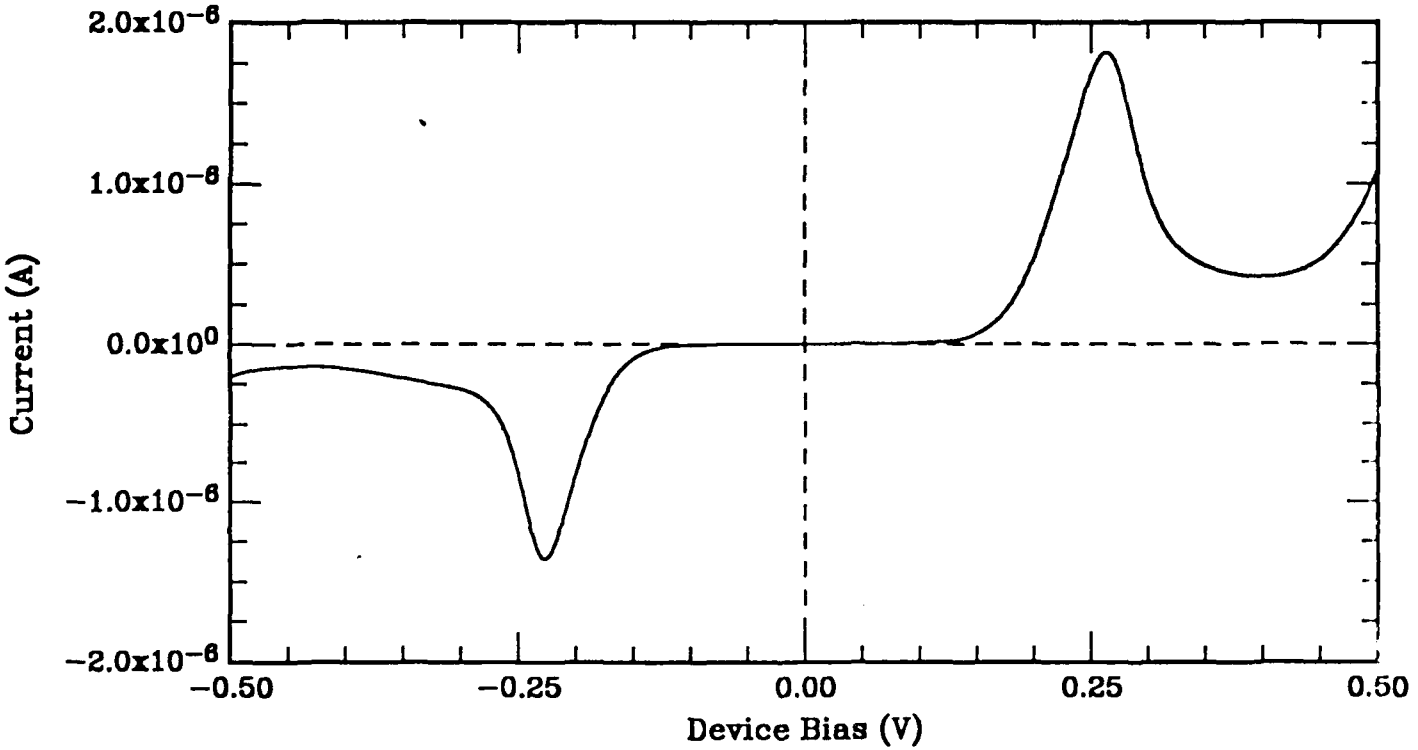
Figures Captions

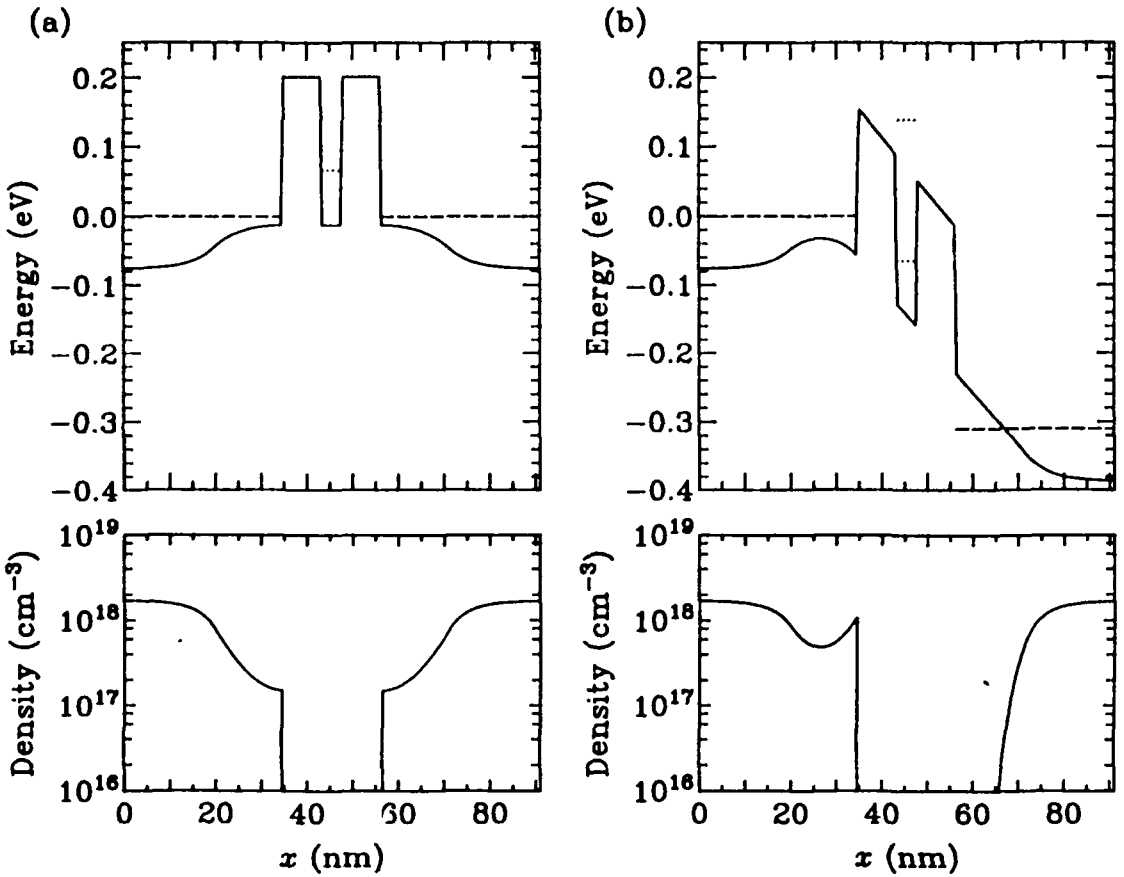
Figure 1. Self-consistent band diagram using Poisson's equations for the electrostatic potential. The electrons in the contacts are treated in a finite-temperature Thomas-Fermi approximation. The simulation does not include current-flow. The structure is a 85Å $Al_xGa_{1-x}As$ (x = .264) barrier / 44Å GaAs QW / 85Å $Al_xGa_{1-x}As$ (x = .264) barrier structure at T = 77K for (a) no applied bias and (b) resonant bias. The energies of the bound states are denoted by dashed lines and the Fermi level by a dotted line.

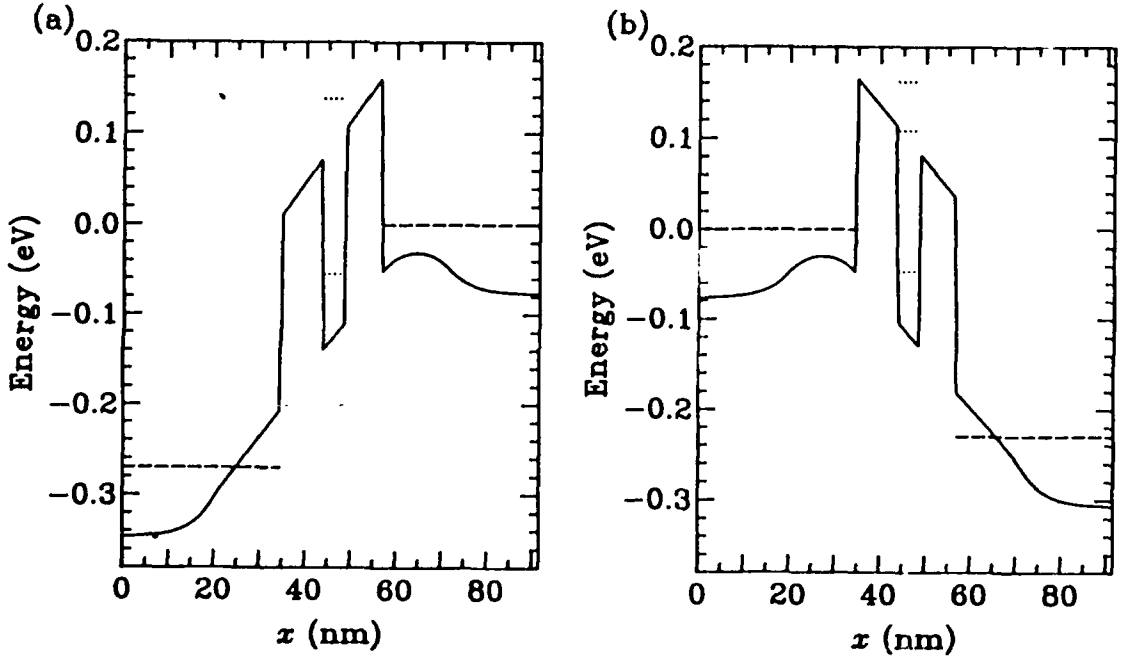
Figure 2. Current voltage characteristics of the 85Å sample for square mesa areas of 1.6×10^{-7} cm^2. Positive voltage corresponds to electron injection from the top contact. T = 77K.

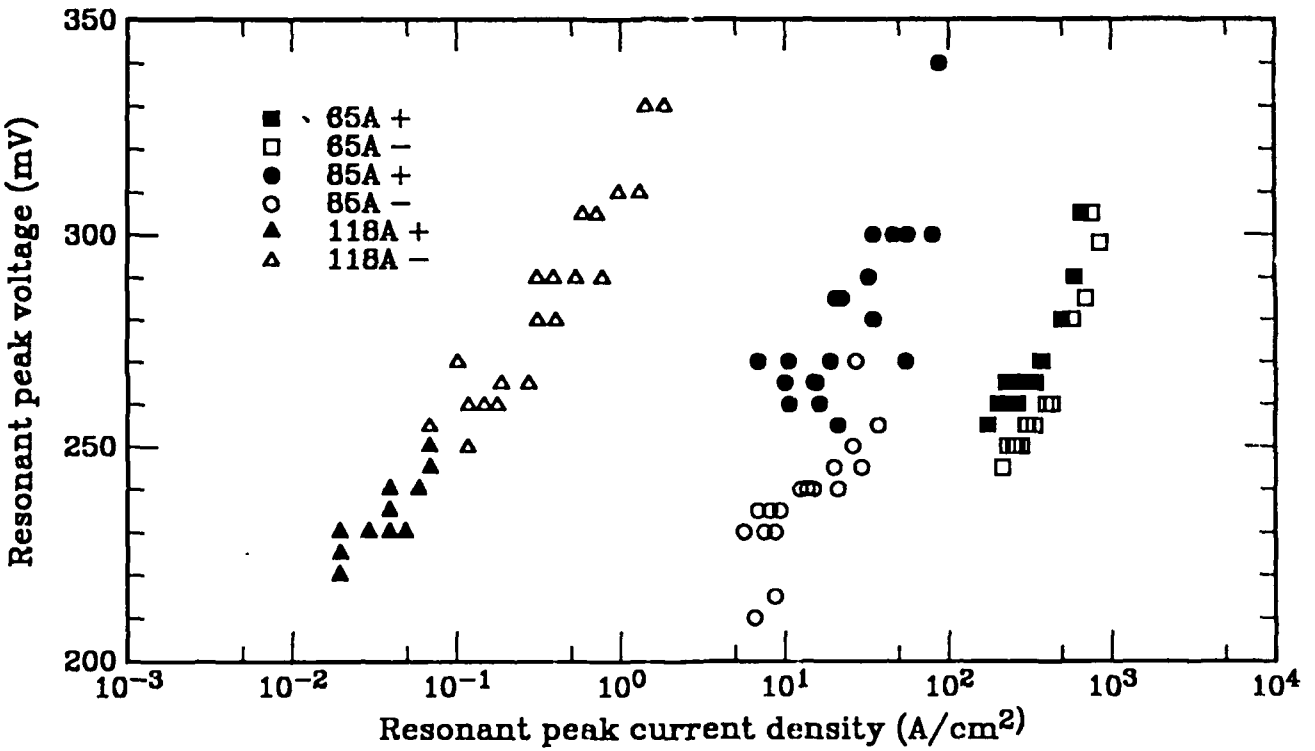
Figure 3. Resonant peak voltage position versus resonant peak current density for a large sample of three different barrier thickness structures. Both positive (+) and negative (-) voltage polarities are shown for the three structures of nominal barrier thickness 65Å, 85Å, and 118Å. T = 77K.

Figure 4. Self-consistent band diagrams of a 90Å bottom barrier / 48Å quantum well / 80Å top barrier structure at resonance. T = 77K. (a) Positive bias and (b) negative bias polarities are the same as Figure 2.


2.0x10⁻⁶
1.0x10⁻⁶
0.0x10⁰
−1.0x10⁻⁶
−2.0x10⁻⁶
Current (A)
−0.50
−0.25
0.00
0.25
0.50
Device Bias (V)


(a)
(b)
Energy (eV)
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
Density (cm^{-3})
10^{19}
10^{18}
10^{17}
10^{16}
0
20
40
60
80
x (nm)


(a)
(b)
Energy (eV)
0.2
0.1
0.0
−0.1
−0.2
−0.3
0
20
40
60
80
x (nm)


350
300
250
200
Resonant peak voltage (mV)
65A +
65A −
85A +
85A −
118A +
118A −
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
Resonant peak current density (A/cm^2)